

IOWA STATE UNIVERSITY

Digital Repository

Retrospective Theses and Dissertations


Iowa State University Capstones, Theses and
Dissertations

1989

The bootstrap method for Markov chains

Cheng-Der Fuh
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>

 Part of the [Mathematics Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Fuh, Cheng-Der, "The bootstrap method for Markov chains " (1989). *Retrospective Theses and Dissertations*. 9038.
<https://lib.dr.iastate.edu/rtd/9038>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9003517

The bootstrap method for Markov chains

Fuh, Cheng-Der, Ph.D.

Iowa State University, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

The bootstrap method for Markov chains

by

Cheng-Der Fuh

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Co-majors: Mathematics
Statistics

Approved _____ Signature was redacted for privacy.

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Departments

Signature was redacted for privacy.

For ~~the~~ Graduate College

Iowa State University
Ames, Iowa
1989

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
1 INTRODUCTION AND PRELIMINARIES	1
1.1 Overview	1
1.2 The Bootstrap Method	6
1.3 The General State Space Markov Chain	10
2 BOOTSTRAPPING A FINITE STATE MARKOV CHAIN	13
2.1 Introduction	13
2.2 Central Limit Theorem for the Transition Probability Estimator	16
2.3 Bootstrapping the Transition Probability of a Markov Chain	22
2.4 Bootstrapping the Hitting Time of a Markov Chain	32
3 BOOTSTRAPPING AN INFINITE STATE MARKOV CHAIN	37
3.1 Introduction	37
3.2 Estimator I	39
3.3 Estimator II	43
3.4 Estimator III	48
4 LIMIT THEOREMS FOR AN ARRAY OF HARRIS CHAINS	52

4.1	Central Limit Theorem for an Array of Harris Chains	52
4.2	Applications and Further Research	59
5	BIBLIOGRAPHY	66

ACKNOWLEDGEMENTS

I wish to acknowledge the expert guidance of Dr. Krishna Athreya during my graduate studies at Iowa State University. I not only learned a great deal from him, I also enjoyed the learning. The lesson I really appreciate is that one can have fun proving results in probability theory.

I also wish to thank Dr. Dean Isaacson for his encouragement and support during my graduate studies at Iowa State University.

1 INTRODUCTION AND PRELIMINARIES

1.1 Overview

The topic of inference for Markov chains has received much attention in the sixties, and the work culminated in a comprehensive book by Billingsley [12]. In recent years several publications on the topic have appeared on its special aspects. One of these problems is to estimate the distribution of the first hitting time of a given state. Such problems arise in several areas of applied probability, e.g., queueing theory and reliability. But even for a simple stochastic model like a finite state Markov chain, the computation for the distribution of the maximum likelihood estimator of the hitting time is rather complicated. In more precise formulation, the problem is as follows:

Let $\{X_n; n \geq 0\}$ be a homogeneous ergodic (positive recurrent, irreducible and aperiodic) Markov chain, with transition probability matrix $P = (p_{ij})$ and countable state space S . The ergodic property of the Markov chain leads to the existence of the unique invariant probability measure $\Pi \equiv \{\pi_j\}$, which is determined by the balance equation:

$$\pi_j = \sum_i \pi_i p_{ij}.$$

Let $\mathbf{x} = \{x_0, x_1, \dots, x_n\}$ be a realization of the process observed up to time n . From Derman [16], we know that the maximum likelihood estimator

$\hat{P}_n \equiv (\hat{p}_n(i, j))$ of P is given by:

$$\hat{p}_n(i, j) = \begin{cases} n_{ij}/n_i, & \text{if } n_i > 0; \\ \delta_{ij}, & \text{otherwise,} \end{cases} \quad (1.1)$$

where

$$\begin{aligned} n_{ij} &= \text{number of } ij \text{ transitions observed up to time } n, \\ n_i &= \text{number of visits to state } i \text{ observed up to time } n. \end{aligned}$$

Since $\Pi = \Pi(P)$ is determined by P , a natural estimator for Π is

$$\Pi(\hat{P}_n) \equiv \hat{\Pi}_n \equiv (\hat{\pi}_n(i)),$$

where

$$\hat{\pi}_n(i) = \frac{n_i}{n}. \quad (1.2)$$

Without loss of generality, we may assume that the initial state for the Markov chain is 1, that is, $x_0 = 1$. Let T_Δ be the first hitting time of the given state Δ , that is,

$$T_\Delta = \begin{cases} \inf\{n; n \geq 0, X_n = \Delta\}; \\ \infty, & \text{if no such } n \text{ exist.} \end{cases}$$

Let $Pr(t; P) \equiv P\{T_\Delta \leq t \mid X_0 = 1; P\}$, denote the probability that $T_\Delta \leq t$, for $t \in \{1, 2, \dots\}$, for a Markov chain $\{X_n\}$ with transition probability P and initial state $x_0 = 1$. Even in the simple case when P is a finite matrix, which occurs quite often in practice, it is difficult to compute $Pr(t; P)$ or to find the expected value of T_Δ . However, the method of bootstrap could prove useful here.

The bootstrap method for estimating the distribution of a pivotal quantity of an estimator was originally proposed by Efron [19]. This can be described as follows:

1) Let $R(\underline{X}, F)$ be a random variable of interest, where $\underline{X} = (X_0, X_1, \dots, X_n)$ indicates the entire sample and F is the unknown distribution. On the basis of having observed $\underline{X} = \underline{x}$, we wish to estimate some aspect of R 's distribution.

2) Construct \hat{F}_n , an estimate of the probability distribution F , based on the observed realization \underline{x} .

3) With the original sample fixed, draw a "bootstrap sample" of size m , from a population with distribution function \hat{F}_n . Denote this sample by

$$\underline{X}^* = (X_0^*, X_1^*, \dots, X_m^*).$$

4) Approximate the sampling distribution of $R(\underline{X}, F)$ by the bootstrap distribution of $R^* \equiv R(\underline{X}^*, \hat{F}_n)$, which can be calculated by Monte Carlo simulation that generates a large number of bootstrap samples.

Efron [20] has shown that for a large number of statistics of interest, the distribution of R^* approximates that of R under some regularity conditions. Several authors have extended these results. The asymptotic theory for the bootstrap estimator, gives on the one hand some guidelines for the practical use of this method, and on the other enriches the subject of limit theorems in probability theory. All of these will be reviewed in Section 1.2.

The application of the bootstrap method to estimate the distribution of hitting time T_Δ of a given state Δ in the homogeneous ergodic Markov chain was originated by Kulperger and Prakasa Rao [30]. The basic question here is to verify that the pivotal quantity $\sqrt{n}(\hat{p}_n(i, j) - p_{ij})$ of $\hat{p}_n(i, j)$ has the same asymptotic distribution as the pivotal quantity $\sqrt{N_n}(\tilde{p}_n(i, j) - \hat{p}_n(i, j))$ of $\tilde{p}_n(i, j)$, where $\tilde{p}_n(i, j)$ is the

bootstrap estimator of p_{ij} and N_n is the bootstrap sample size. This has been done by Kulperger and Prakasa Rao [30] for the finite state Markov chain case. A different approach to these problems will be taken in Chapter 2.

The classical central limit theorem for the estimator \hat{P}_n of the transition probability matrix P , has been proved by Billingsley [12] under some regularity properties.

Theorem 1 *Let X be a homogeneous ergodic Markov chain with finite state space and transition probability matrix P , let \hat{P}_n be the maximum likelihood estimator of P defined as above. Then*

$$\sqrt{n}(\hat{P}_n - P) \mapsto N(Q, \Sigma_P) \text{ in distribution,}$$

where Σ_P is the variance-covariance matrix which is continuous as a function of P with respect to the supremum norm on the class of $k \times k$ stochastic matrices.

In Section 2.2 we will give a proof of the above result based on the idea that the evolution of a recurrent Markov chain can be broken up into independent and identically distributed cycles. These will be defined in Chapter 2. With this setting and Lindeberg's central limit theorem, we prove the asymptotic normality of the "bootstrap estimator" under some suitable conditions in Section 2.3. The bootstrap method for the hitting time and the expected value of the hitting time are in Section 2.4.

The generalization from finite state Markov chains to infinite state Markov chains is in Chapter 3. In Section 3.2, we generalize the bootstrap method which we used in Chapter 2. The bootstrap algorithm works well under some suitable conditions for Markov chains. The relation between the original sample size n and the bootstrap resample size N_n , which relates to the problem of ergodic coefficient in Markov chain,

is somewhat complicated and needs to be studied. But this is not a serious problem for the method of bootstrap. Due to the availability of cheap high speed computation, we can generate as many bootstrap samples as we wish at relatively low cost.

Successive excursions from a fixed state Δ in an ergodic Markov chain are i.i.d. This is the device of the well known regeneration method for a Markov chain. By using this idea, we have a new method to generate the bootstrap sample, that is, we decompose the entire sequence $\{X_0, X_1, X_2, \dots, X_n, \dots\}$ as $\{\eta_j; j = 0, 1, \dots\}$ where $\eta_j \equiv \{X_{T_\Delta^{(j)}}, \dots, X_{T_\Delta^{(j+1)} - 1}\}$ and $\{T_\Delta^{(j)}; j = 0, 1, 2, \dots\}$ are the successive times for a recurrent state Δ . By the strong Markov property, it is clear that $\{\eta_j, j = 1, 2, 3, \dots\}$ forms an i.i.d. sequence. Therefore, the bootstrap method for the i.i.d. case can be used here. If the sample size n is fixed, then the number of full cycles k that are included in $\{X_0, \dots, X_n\}$ is random. On the other hand, if we observe the process until k full cycles are obtained, then the sample size is random. This leads to two different bootstrap resample methods.

The asymptotic properties of the bootstrap method in the i.i.d. case have been investigated by Athreya [3] [5], Bickel and Freedman [11], and Singh [38] and others. Their results will prove useful in the studies of the asymptotic properties of the bootstrap methods for the Markov chain case. All of these three different bootstrap methods and their corresponding asymptotic properties are discussed in Chapter 3.

In Chapter 4, we consider the central limit theorem for a double array of Harris chains. The application of these theorems to the bootstrap estimator and further research are in Section 4.2. A brief review of the theory of Harris chains is given in Section 1.3 below.

1.2 The Bootstrap Method

Let X_1, X_2, \dots, X_n be independent, identically distributed random variables with distribution function F . Suppose $R(\underline{X}, F)$ is a random variable of both the observation \underline{X} , and the distribution F . The bootstrap method introduced by Efron [19], is designed to estimate the sampling distribution of $R(\underline{X}, F)$ on the basis of the observed data \underline{x} . This will be particularly useful either when F is unknown or when F is known but $R(\underline{X}, F)$ has a complicated distribution.

The difficult part of the bootstrap procedure is the actual calculation of the bootstrap distribution. Three methods of calculation have been suggested by Efron [19]:

Method 1) Direct theoretical calculation. This is almost impossible except for some extremely simple cases.

Method 2) Taylor series expansion methods can be used to obtain the approximate mean and variance of the bootstrap distribution of R^* .

Method 3) Monte Carlo approximation to the bootstrap distribution. Repeated realizations of \underline{X}^* are generated by taking random samples of size n from \hat{F}_n , say $\underline{x}^{*1}, \underline{x}^{*2}, \dots, \underline{x}^{*N}$, and the histogram of the corresponding values $R(\underline{x}^{*1}, \hat{F}_n), R(\underline{x}^{*2}, \hat{F}_n), \dots, R(\underline{x}^{*N}, \hat{F}_n)$ is taken as an approximation to the actual bootstrap distribution.

Due to the increased availability of high speed computing, Method 3) is the most commonly used and has led to the extensive use of the bootstrap technique in many branches of applied statistics. The reader is referred to Efron [19] [20], Freedman [22], Freedman and Peters [23], and Wu [39] for the details.

Some asymptotic theory for the bootstrap has become available in the literature,

for example, see Babu and Singh [8], Beran [10], Bickel and Freedman [11], Hall [25] [26] [27] and Singh [38]. Most of the effort in these papers is devoted to showing that the bootstrap distribution of many important statistics is asymptotically the same as that of the original statistic itself.

So far, most of the bootstrap methods have been used for the i.i.d. case. The basic reference is Efron [19]. The comparison between the bootstrap method and the weighted jackknife method is in Wu [39], and Singh [38] and Hall [26] discuss Edgeworth expansions for the bootstrap. The investigation of statistical inference problems on stochastic processes is usually more complicated than the classical i.i.d. case. The bootstrap algorithm should prove to be a useful technique for these cases. Bose [13] discussed the bootstrap method for the autoregression model. The paper by Kulperger and Prakasa Rao [30] deals with finite state Markov chain problems.

It is now known that the bootstrap technique is not “the man of all seasons”, or “the remedy of all diseases”. There are situations where bootstrap methods lead to incorrect conclusions. Some basic counterexamples are in Efron [19], Bickel and Freedman [11], and Wu [39]. The failure of the bootstrap of the mean in the case of heavy tails is discussed by Athreya [5] [6]. It has been shown that the bootstrap is not consistent for estimating the distribution of the mean when the original population is from the domain of attraction of a non-normal stable law. In this case, the limiting distributions of the sample mean and its bootstrap version are quite different, the latter one being a random probability distribution. Similarly, for variance estimation using naive bootstrap could be bad if the underlying population has no fourth moment. There are some modifications of the bootstrap method such as changing the resample size from n to m with $m = o(n)$ or trimming the sample and doing

bootstrap on the reduced sample. Arcones and Gine [1], have shown that when X is in the domain of attraction of a stable law of order p and $m(\log \log n)/n \mapsto 0$, then the bootstrap CLT holds a.s. The idea of a bootstrap resample size different from the original sample size will be used in this paper, especially, for the bootstrap estimator of the transition probability matrix in an ergodic Markov chain with infinite state space.

The necessary conditions for the bootstrap of the mean to work are given by Hall [27], and Gine and Zinn [24]. They show that the bootstrap distribution function of the mean, suitably normalized, converges in probability to some fixed nondegenerate distribution function if, and only if, either (a) the original distribution is in the domain of attraction of the normal law; or (b) the original distribution has slowly varying tails and one of the two tails completely dominates the other. In case (a) the limiting distribution is normal. In case (b) it is Poisson with unit mean. Only case (a) is statistically interesting, since the limiting distributions of the sample mean and its bootstrap version coincide. Therefore, the bootstrap is (weakly) consistent if, and only if, the sampling distribution is from the domain of attraction of the normal law.

Let us consider some examples before completing this section:

Example 1 (*The average*)

Suppose $\underline{X} = \{X_i; i = 1, 2, \dots, n\}$ are i.i.d. random variables with unknown distribution F . Let $R(\underline{X}, F) = \sqrt{n} \frac{(\bar{X}_n - \mu)}{\sigma}$, where $\mu = EX_1$, $\sigma^2 = \text{Var}(X_1)$, and \bar{X}_n is the mean of \underline{X} .

Let

$$J_n(x, F) = P_F\{R(\underline{X}, F) \leq x\}.$$

Then the bootstrap estimator of $J_n(\cdot, F)$ is $J_n(\cdot, \hat{F}_n)$. With $EX^2 < \infty$, we have

$$\sup_x |J_n(x, \hat{F}_n) - J_n(x, F)| \xrightarrow{w.p.1} 0$$

Since \hat{F}_n is a discrete distribution, it is possible, in principle, to calculate $J_n(x, \hat{F}_n)$ as follows:

Let $\{(X_{j1}^*, \dots, X_{jn}^*); 1 \leq j \leq M\}$ be M pseudo-random samples of size n drawn (with replacement) from the empirical cumulative distribution function \hat{F}_n .

Let $\bar{X}_{jn}^* = \frac{1}{n} \sum_{t=1}^n X_{jt}^*$ be the mean of the j^{th} pseudo-random sample. Then

$$J_{nM}^*(x) = \frac{1}{M} \sum_{j=1}^M I[\sqrt{n}(\frac{\bar{X}_{jn}^* - \bar{X}_n}{s}) \leq x]$$

is the Monte Carlo approximation to $J_n(x, \hat{F}_n)$, where s is the sample standard deviation.

Example 2 (Kolmogorov-Smirnov Statistic)

Suppose $\underline{X} = \{X_i; i = 1, 2, \dots, n\}$ are i.i.d. random variables with unknown distribution function F . Let $R(\underline{X}, F) = \sqrt{n} \sup_x |\hat{F}_n(x) - F(x)|$, where $\hat{F}_n(\cdot)$ is the empirical c.d.f. based on \underline{X} . Let

$$J_n(x, F) = P_F\{R(\underline{X}, F) \leq x\}.$$

Then the bootstrap estimator of $J_n(\cdot, F)$ is $J_n(\cdot, \hat{F}_n)$. From Beran [10], we have

$$\sup_x |J_n(x, \hat{F}_n) - J_n(x, F)| \xrightarrow{w.p.1} 0$$

Since \hat{F}_n is a discrete distribution, it is possible, in principle, to calculate $J_n(x, \hat{F}_n)$ as follows:

Let $\{(X_{j1}^*, \dots, X_{jn}^*); 1 \leq j \leq M\}$ be M pseudo-random samples of size n drawn (with replacement) from the empirical cumulative distribution function \hat{F}_n .

Let $F_{jn}^*(x) = \frac{1}{n} \sum_{i=1}^n I(X_{ji}^* \leq x)$ be the empirical cumulative distribution function of the j^{th} pseudo-random sample. Then

$$J_{nM}^*(x) = \frac{1}{M} \sum_{j=1}^M I[\sqrt{n} \sup | F_{jn}^*(t) - \hat{F}_n(t) | \leq x]$$

is the Monte Carlo approximation to $J_n(x, \hat{F}_n)$.

1.3 The General State Space Markov Chain

The theory of Markov chains on a general state space was originated by Doeblin [18]. Significant contributions to its development have been made by Harris [28], Orey [34], Nummelin [33], and Athreya and Ney [7]. The classical approach can be found in the books written by Neveu [31], Orey [34], and Revuz [35]. A new approach using the device of an embedded renewal sequence is in Nummelin's book [33].

For an irreducible Markov chain, a point x_0 is recurrent if for any initial point x , it returns to x_0 with probability one. Such chains can be studied by using the embedded renewal process of returns to x_0 . This covers the countable state space case. In the general state space, such a point may not exist. And this makes the theory for this case a little difficult. Doeblin [18] proved an ergodic theorem in this case under a strong hypothesis, known as Doeblin's condition, namely, there exists a probability measure φ on the state space (S, \mathcal{S}) , numbers $\epsilon > 0$, $\delta < 1$, and an integer $n_0 < \infty$, such that for all $x_0 \in S$, $E \in \mathcal{S}$ we have $P^{n_0}(x_0, E) > \delta$, whenever $\varphi(E) \geq \epsilon$.

Harris [28] introduced a weaker condition as follows:

Definition 1 A Markov chain X_n is Harris recurrent or φ -recurrent if there exists a σ -finite measure φ on the state space (S, \mathcal{S}) such that $P_x(X_n \in A \text{ for some } n) = 1$ for all $A \in \mathcal{S}$ with $\varphi(A) > 0$, where P_x denote the probability starting at x .

A definition equivalent to the above was given by Athreya and Ney [7] as follows:

Definition 2 $\{X_n\}$ is $(A, \lambda, \varphi, n_0)$ recurrent if there exists a set $A \in \mathcal{S}$, a probability measure φ on A , a number $\lambda \in (0, \infty)$, and an integer $n_0 \in (0, \infty)$ such that

- 1) $P_x\{X_n \in A \text{ for some } n \geq 1\} = 1$ for all $x \in S$,
- 2) $P_x\{X_{n_0} \in E\} \geq \lambda\varphi(E)$ for all $x \in A$ and $E \subset A$.

Here A is called a regeneration set. For example, this holds for a one-point set $A = \{x\}$ if, and only if, x is a recurrent state, since we may take an arbitrary $n_0 > 0$, $\lambda = 1/2$ and $\varphi(E) = P^{n_0}(x, E)$. The following example is a typical application and shows that regeneration sets exist in far more general situations:

Example 3 Assume that the transition functions contain components with smooth densities. That is, for some μ and n_0 , we have

$$P^{n_0}(x, E) \geq \int_E f^{n_0}(x, y) \mu(dy),$$

$$S_0 = \{x \in S : \int f^{n_0}(x, y) \mu(dy) > 0\} \text{ is not empty,}$$

where $f^{n_0}(x, y)$ is jointly continuous in (x, y) in a suitable topology on S . Then a regeneration set exists, provided that for some $x_0 \in S_0$, every neighborhood of x_0 is recurrent. Indeed, choose $y_0 \in \text{supp } \mu$ with $\delta = f^{n_0}(x_0, y_0) > 0$ and let R_x, R_y be neighborhoods of x_0, y_0 with $f^{n_0}(x, y) \geq \delta/2, x \in R_x, y \in R_y$. Then if $\varphi(E) = \mu(E \cap R_y)/\mu(R_y)$, we have for $x \in R_x$ that

$$P^{n_0}(x, E) \geq \int_{E \cap R_y} f^{n_0}(x, y) \mu(dy) \geq \frac{\delta}{2} \mu(E \cap R_y) = \frac{\delta \mu(R_y)}{2} \varphi(E).$$

We shall justify the term “regeneration set” by showing that it is possible to construct $\{X_n\}$ simultaneously with a renewal process T_1, T_2, \dots with respect to which the Markov chain becomes regenerative. The method uses a randomization technique. The reader is referred to Athreya and Ney [7] for the details.

The regeneration points obviously behave rather like stopping times, but are not so in the strict sense, since in addition to $\mathcal{F}_\infty = \sigma(X_t; t \in T)$, they also depend on the 0 – 1 variables determining the randomization. However, they are in the category of so called *randomized stopping times*. We will not go into a discussion of this topic here, and refer to Nummelin [33] for the details.

The equivalence between the above two definitions is well known. We remark that in practical cases, the second definition seems easier to check than the first one. For example, for $S = R$, the obvious choice of φ is frequently Lebesgue measure (possibly restricted to some interval), and it may be fairly easy to check that every interval is recurrent. But to check the first definition, one needs to show recurrence of every Borel set of positive Lebesgue measure, and since such a set A can have a very complicated structure (e.g., A need not have interior points), this could prove to be a difficult task.

2 BOOTSTRAPPING A FINITE STATE MARKOV CHAIN

2.1 Introduction

We consider a Markov chain $\{X_0, X_1, \dots\}$ with discrete (i.e., finite or countable) state space $S = \{1, 2, 3, \dots\}$ and transition probability matrix $P = (p_{ij})$, $i, j \in S$. By this we mean that P is a given $|S| \times |S|$ matrix such that $p_{i\cdot} = (p_{ij})$, $j \in S$ is a probability (vector) for each i , and that we study $\{X_n\}$ subject to exactly those governing probability laws $\mathcal{P} = \mathcal{P}_\mu$ (Markov probabilities) for which

$$\mathcal{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mu(i_0)p_{i_0 i_1}p_{i_1 i_2} \cdots p_{i_{n-1} i_n},$$

where $\mu(i) \equiv \mathcal{P}(X_0 = i)$. It is well known that with the fixed transition probability P , there is one-to-one correspondence between the Markov probabilities \mathcal{P} and the set of initial distributions.

If \mathcal{P} is a Markov probability, then (with the usual a.s. interpretation of conditional probabilities and expectations)

$$p_{ij} = \mathcal{P}_i(X_1 = j) = \mathcal{P}(X_{n+1} = j \mid X_n = i), \quad (2.1)$$

$$\mathcal{P}(X_{n+1} = j \mid \mathcal{F}_n) = p_{X_n j} = \mathcal{P}_{X_n}(X_1 = j), \quad (2.2)$$

where $\mathcal{F}_n \equiv \sigma(X_0, X_1, \dots, X_n)$, the σ -algebra generated by $\{X_0, X_1, \dots, X_n\}$. Conversely, (2.2) is a sufficient condition for \mathcal{P} to be a Markov probability. The

intuitive contents of (2.2) is that the chain can be constructed by, at step n , drawing X_{n+1} according to P_{X_n} (to get started, draw X_0 according to μ).

In practice, often one does not know the transition probability P . So, it is of interest to estimate or test hypotheses about P on the basis of a set of observations $\mathbf{x} = \{x_0, x_1, x_2, \dots, x_n\}$ of an initial segment $\{X_0, X_1, X_2, \dots, X_n\}$. The maximum likelihood estimator defined in Section 1.1 is a natural estimator.

We shall only consider ergodic Markov chains on discrete state space S , for which there exists a unique stationary distribution $\Pi \equiv \{\pi_j; j = 1, 2, \dots, k\}$ determined by balance equations

$$\pi_j = \sum_i \pi_i p_{ij}, \quad j = 1, 2, \dots, k. \quad (2.3)$$

For a subset of states $A \subset S$, the first hitting time T_A is

$$T_A = \min\{t > 0 : X_t \in A\}.$$

Positive recurrence implies $E_i T_A < \infty$; here $E_i(\cdot)$, $P_i(\cdot)$ denote expectation and probability given $X_0 = i$ respectively. The mean hitting times are determined by an elementary equation. For fixed A does not contain i , the means $h(i) = E_i T_A$ satisfy

$$h(i) = 1 + \sum_j p_{ij} h(j). \quad (2.4)$$

In practical examples there is often special structure which enables us to find the stationary distribution Π without solving equations (2.3) (e.g., reversibility; double-stochasticity), whereas there are no such useful tricks solving (2.4) for hitting times. This leads to finding an estimator of the transition probability P , and the distribution of the estimator. The purpose of this chapter is to show that a new technique called bootstrap, which was proposed by Efron [19] in 1979 in another context, can be used

to estimate the distribution of the estimator of the transition probability P . The asymptotic normality of \hat{P}_n , the maximum likelihood estimator of P , is well known, and hence can be used to approximate the distribution of \hat{P}_n . The advantage of using the bootstrap method is in part to avoid the complexity of computation of the variance of \hat{P}_n .

Let us now discuss some examples which indicate the importance of first hitting time to a given state in a Markov chain.

Example 4 *Card-shuffling.*

Repeated shuffling of an N – card deck can be modeled by a Markov chain whose states are the $N!$ possible configuration of the deck and whose transition matrix depends on the method for doing a shuffle. Then, we get a doubly-stochastic chain (in fact, a random walk on the permutation group). Details can be found in Diaconis[17]. Here $\pi(i) = 1/N!$ for each configuration i . Consider the number of shuffles T needed until a particular configuration i is reached. The problem here is to estimate the first hitting time T .

Example 5 *A simple reliability model.*

Consider a system with K components. Suppose components fail and are repaired, independently for different components. Suppose component i fails at exponential rate a_i and is repaired at exponential rate b_i . Then the process evolves as a Markov chain whose states are subsets $B \subset \{1, 2, \dots, k\}$ representing the set of failed components. There is some set \mathcal{F} of subsets B which imply system failures, and the problem here is to estimate the time $T_{\mathcal{F}}$ until system failure.

Example 6 *Basic single server queue.*

Here the states are $\{0, 1, 2, \dots\}$, $Q(i, i+1) = a$, $Q(i, i-1) = b$ for $i \geq 1$; the parameters a and b represent the arrival and service rates; and $a < b$ for stability. The stationary distribution is geometric: $\pi(i) = (1 - a/b)(a/b)^i$. The problem here is to estimate T_K , the time until the queue length first reaches K , where K is sufficiently large such that $\pi[K, \infty) = (a/b)^K$ is small; that is, the queue length rarely exceeds K .

Example 7 *Two M/M/1 queues in series.*

Suppose each server has service rate b , and let the arrival rate be $a < b$. Write (X_t^1, X_t^2) for the queue lengths process. It is well known that the stationary distribution (X^1, X^2) has independent geometric components:

$$\pi(i, j) = (1 - a/b)^2 (a/b)^{i+j}; \quad i, j \geq 0.$$

The problem here is to estimate T_K , the time until the combined length $X_t^1 + X_t^2$ first reaches k .

Here there are two different kinds of state spaces S for the above given Markov chain models. In the first two examples, it is a finite state space, but it is infinite for the next two examples. The bootstrap method for ergodic Markov chains will be discussed based on the state space. The finite state space case is in this chapter, while the infinite state space case will be in Chapter 3.

2.2 Central Limit Theorem for the Transition Probability Estimator

Let $\{X_n; n \geq 0\}$ be a homogeneous ergodic Markov chain with finite state space $S = \{1, 2, \dots, k\}$ and transition probability matrix $P = (p_{ij})$. Without loss of

generality, we may assume that p_{ij} are positive for all $i, j \in S$, since by irreducibility there exists $N \geq 1$ such that $P^N = (p_{ij}^N)$ has all entries positive. This implies that there exists an invariant probability measure $\Pi = (\pi_1, \dots, \pi_k)$ such that $\pi_j > 0$, $\sum_{j=1}^k \pi_j = 1$, $p_{ij}^{(n)} \mapsto \pi_j$, as $n \rightarrow \infty$ for all $i \in S$ and $\pi_j = \sum_{\alpha} \pi_{\alpha} p_{\alpha j}$, $j = 1, \dots, k$.

Suppose $\mathbf{x} = \{x_0, x_1, \dots, x_n\}$ is a realization of the process $\{X_j; j = 0, \dots, n\}$ observed up to time n . We estimate P by its maximum likelihood estimator $\hat{P}_n \equiv (\hat{p}_n(i, j))$, where

$$\hat{p}_n(i, j) = \begin{cases} n_{ij}/n_i, & \text{if } n_i > 0; \\ \delta_{ij}, & \text{otherwise,} \end{cases} \quad (2.5)$$

and estimate Π by $\hat{\Pi}_n \equiv (\hat{\pi}_n(i, j))$, where

$$\hat{\pi}_n(i, j) = \frac{n_i}{n}, \quad (2.6)$$

n_{ij} = observed number of ij transitions in $\{x_0, \dots, x_n\}$,

n_i = observed number of visits to state i in $\{x_0, \dots, x_n\}$.

Since the state space S is finite, we can consider the non-parametric case as a special case of the parametric case. So, the consistency and asymptotic normality of the maximum likelihood estimators can be deduced using the analogy with the multinomial distribution. This idea also can be used to prove the bootstrap estimators of \hat{P}_n given \mathbf{x} .

The consistency of $\hat{\Pi}_n$ for Π follows from the strong law applied to the renewal sequence of return times to state i .

Theorem 2 *Let $X = \{X_n; n \geq 0\}$ be a homogeneous ergodic Markov chain, then for all i*

$$\hat{\pi}_n(i) = n_i/n \mapsto \pi_i \quad \text{with probability 1} \quad \text{as } n \rightarrow \infty.$$

Before we give the statement and proof of the asymptotic normality of the maximum likelihood estimator \hat{P}_n , we state the famous Kolmogorov's inequality and the Cramér-Wold device as lemmas to be used in Theorem 3.

Lemma 1 (*Kolmogorov's inequality*) *Let $\{X_n\}$ be independent random variables such that*

$$E(X_n) = 0, \quad E(X_n^2) = \sigma^2(X_n) < \infty \quad \text{for all } n.$$

Then for every $\epsilon > 0$, we have

$$P\{\max_{1 \leq j \leq n} |S_j| > \epsilon\} \leq \frac{\sigma^2(S_n)}{\epsilon^2},$$

where $S_j = X_1 + \cdots + X_j$, $j = 1, 2, \dots, n$.

Lemma 2 (*Cramér-Wold device*) *Let*

$$X_n = (X_{n1}, \dots, X_{nk}) \text{ and } X = (X_1, \dots, X_k) \text{ be random vectors in } R^k.$$

Then, X_n converges in distribution to X if, and only if, each linear combination of the components of X_n converges in distribution to the corresponding linear combination of the components of X .

Theorem 3 *Let $X \equiv \{X_n; n \geq 0\}$ be a homogeneous ergodic Markov chain, and \hat{P}_n the maximum likelihood estimator of P defined in (2.5), Then*

$$\sqrt{n}(\hat{P}_n - P) \mapsto N(0, \Sigma_P) \text{ in distribution,}$$

where Σ_P the variance-covariance matrix which is a $k^2 \times k^2$ block diagonal matrix and is continuous as a function of P with respect to the supremum norm on the class of $k \times k$ stochastic matrices.

Proof. The Markov chain $\{X_n\}$ can be viewed as having been generated in the following fashion: Consider an independent collection of random variables $\{W_{it}\}$, $i = 1, \dots, k$, $t = 1, 2, \dots$ such that for each t , $P\{W_{it} = j\} = p_{ij}$, $i = 1, \dots, k$, $j = 1, \dots, k$. Now, let

$$\begin{aligned} X_0 &= 1, \\ X_{t+1} &= W_{X_t m}, \end{aligned}$$

where $m - 1$ is the number of l with $1 \leq l \leq n$ such that $X_l = X_t$.

Now, for fixed i , define

$$C_t(j) = \begin{cases} 1, & \text{if } W_{it} = j, \ t = 1, 2, \dots, n_i; \\ 0, & \text{otherwise.} \end{cases}$$

By the definition given above, we have

$$\begin{aligned} P\{C_t(j) = 1\} &= p_{ij}, \\ P\{C_t(j) = 0\} &= 1 - p_{ij}. \end{aligned}$$

Note that, n_{ij} defined in (2.6) satisfies

$$n_{ij} = \sum_{t=1}^{n_i} C_t(j).$$

Let

$$m_{ij} \equiv \sum_{t=1}^{[n\pi_i]} C_t(j),$$

where $[x]$ denotes the greatest integer which is less than or equal to x .

Since

$$\frac{n_{ij} - n_i p_{ij}}{\sqrt{n_i}} = \frac{n_{ij} - m_{ij} + m_{ij} - [n\pi_i] p_{ij} + [n\pi_i] p_{ij} - n_i p_{ij}}{\sqrt{n_i}}$$

$$\begin{aligned}
&= \frac{\sqrt{n}}{\sqrt{n_i}} \left(\frac{n_{ij} - n_i p_{ij}}{\sqrt{n}} - \frac{m_{ij} - [n\pi_i] p_{ij}}{\sqrt{n}} \right) \\
&\quad + \sqrt{\frac{[n\pi_i]}{n_i}} \left(\frac{m_{ij} - [n\pi_i] p_{ij}}{\sqrt{[n\pi_i]}} \right),
\end{aligned}$$

and we already have $n_i/n \mapsto \pi_i$ with probability 1 from Theorem 2, it is enough to show that

$$\begin{aligned}
a) \quad & \frac{n_{ij} - n_i p_{ij}}{\sqrt{n}} - \frac{m_{ij} - [n\pi_i] p_{ij}}{\sqrt{n}} \mapsto 0 \text{ in probability,} \\
b) \quad & \left(\frac{m_{ij} - [n\pi_i] p_{ij}}{\sqrt{[n\pi_i]}} \right)_{j=1, \dots, k} \mapsto N(\mathbf{0}, \Gamma_i(P)) \text{ in distribution,}
\end{aligned}$$

where $\Gamma_i(P) = (p_{ij}(\delta_{jl} - p_{il})), \quad j, l = 1, \dots, k$.

We prove a) and b) separately, and prove a) first.

a) Fix i and j , for all $r = 1, 2, 3, \dots$, let

$$\begin{aligned}
e_r &= \begin{cases} 1 - p_{ij}, & \text{if } W_{it} = j, \quad t = 1, 2, \dots, n_i; \\ -p_{ij}, & \text{otherwise.} \end{cases} \\
S_m &= e_1 + \dots + e_m.
\end{aligned}$$

Then,

$$\frac{n_{ij} - n_i p_{ij}}{\sqrt{n}} - \frac{m_{ij} - [n\pi_i] p_{ij}}{\sqrt{n}} = \frac{S_{n_i} - S_{[n\pi_i]}}{\sqrt{n}}.$$

Since n_i/n converges to π_i with probability 1, for any given $\epsilon > 0$, there exists an n_0 such that

$$P\{|n_i - [n\pi_i]| > n\epsilon^3\} < \epsilon \text{ holds for all } n > n_0.$$

For $n > n_0$, we have

$$P\{|S_{[n\pi_i]} - S_{n_i}| > \sqrt{n}\epsilon\}$$

$$\begin{aligned}
&= P\{|n_i - [n\pi_i]| > n\epsilon^3, |S_{[n\pi_i]} - S_{n_i}| > \sqrt{n}\epsilon\} \\
&\quad + P\{|n_i - [n\pi_i]| \leq n\epsilon^3, |S_{[n\pi_i]} - S_{n_i}| > \sqrt{n}\epsilon\} \\
&\leq P\{|n_i - [n\pi_i]| > n\epsilon^3\} \\
&\quad + P\{\max_{|m - [n\pi_i]| \leq n\epsilon^3} |S_m - S_{[n\pi_i]}| > \sqrt{n}\epsilon\} \\
&\leq \epsilon + 2P\{\max_{1 \leq m \leq n\epsilon^3} |S_m| > \sqrt{n}\epsilon\} \\
&\leq \epsilon + 2\frac{1}{n\epsilon^2} \text{Var}(S_{[n\epsilon^3]+1}) \quad (\text{by Lemma 1}) \\
&\leq \epsilon + \frac{2}{n\epsilon^2} ([n\epsilon^3] + 1) \frac{1}{4} \\
&= C\epsilon \quad (\text{where } C > 0 \text{ is a constant}).
\end{aligned}$$

b) By definition, m_{ij} is the sum of an i.i.d. sequence $C_t(j)$ with $E[C_t(j)] = p_{ij}$ and $\text{Cov}(C_t(j), C_t(l)) = p_{ij}\delta_{jl} - p_{ij}p_{il}$. By the classical multi-dimensional central limit theorem for the multinomial case, we have for each i

$$(\frac{m_{ij} - [n\pi_i]p_{ij}}{\sqrt{[n\pi_i]}})_{j=1, \dots, k} \mapsto N(0, \Gamma_i(P)) \text{ in distribution.}$$

Now, for different i , $(m_{ij} - [n\pi_i]p_{ij})/\sqrt{[n\pi_i]}$ are independent. So, by the Cramér-Wold device, We have

$$(\frac{m_{ij} - [n\pi_i]p_{ij}}{\sqrt{[n\pi_i]}})_{i,j=1,2,\dots,k} \mapsto N(0, \Pi_P) \text{ in distribution,} \quad (2.7)$$

where Π_P is a $k^2 \times k^2$ variance covariance matrix with block diagonal form.

Note that $(\frac{n_{ij} - n_i p_{ij}}{\sqrt{n_i}})_{i,j=1,2,\dots,k}$ has the same asymptotic distribution as (2.7).

Since $\frac{n_i}{n} \rightarrow \pi_i$ w.p.1. and

$$\begin{aligned}
\sqrt{n}(\hat{p}_n(i, j) - p_{ij}) &= \sqrt{n}(\frac{n_{ij}}{n_i} - p_{ij}) \\
&= \frac{\sqrt{n}}{\sqrt{n_i}}(\frac{n_{ij} - n_i p_{ij}}{\sqrt{n_i}}),
\end{aligned}$$

it follows that

$$\sqrt{n}(\hat{P}_n - P) \mapsto N(0, \Sigma_P) \text{ in distribution,}$$

where

$$\Sigma_P = \begin{pmatrix} \frac{1}{\pi_1}\Gamma_1(P) & 0 & \cdots & 0 \\ 0 & \frac{1}{\pi_2}\Gamma_2(P) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\pi_k}\Gamma_k(P) \end{pmatrix}. \quad \square$$

The computation of the variance covariance matrix is, in fact, a special case for the general central limit theorem for an array of Harris chains which will be discussed in Section 4.1.

2.3 Bootstrapping the Transition Probability of a Markov Chain

Let $X = \{X_n; n \geq 0\}$ be a homogeneous ergodic Markov chain with finite state space S and transition probability matrix $P = (p_{ij})$. Let $\hat{P}_n = (\hat{p}_n(i, j))$ be the maximum likelihood estimator of P based on the observed data $\underline{x} = (x_0, x_1, \dots, x_n)$ of $X = (X_0, X_1, \dots, X_n)$. The bootstrap method for estimating the sampling distribution of $R(\underline{X}, P) = \sqrt{n}(\hat{P}_n - P)$ can be described as follows:

1) Construct an estimate of the transition probability matrix P , based on the observed realization \underline{x} , (here we use the maximum likelihood estimator \hat{P}_n).

2) With \hat{P}_n fixed, let $X_0^* = x_0$, $X_i^* = x_i^*$, $i = 1, 2, \dots, N_n$, be a realization of a Markov chain with transition probability matrix \hat{P}_n . Call this the bootstrap sample and denote

$$\underline{X}^* = (X_0^*, X_1^*, \dots, X_{N_n}^*),$$

$$\underline{x}^* = (x_0^*, x_1^*, \dots, x_{N_n}^*).$$

3) Approximate the sampling distribution of $R(\underline{X}, P)$ by the bootstrap distribution of $R^* = R(\underline{X}^*, \hat{P}_n)$ conditioned on \underline{X} .

That is, we compute \tilde{P}_n as the maximum likelihood estimator of \hat{P}_n based on \underline{X}^* , and look at the distribution of $\sqrt{N_n}(\tilde{P}_n - \hat{P}_n)$ conditioned on the data $\underline{X} = \underline{x}$. In view of Theorem 3, to justify the bootstrap procedure, we need to prove the asymptotic normality of $\sqrt{N_n}(\tilde{P}_n - \hat{P}_n)$.

Before proving the main theorem in this section, we recall the famous classical Lindeberg-Feller central limit theorem for a double array of random variables. Let us suppose that for every $n \geq 1$, we have a sequence

$$X_{n1}, X_{n2}, \dots, X_{nk_n}$$

of independent random variables with

$$EX_{nj} = 0, \quad V(X_{nj}) = \sigma_{nj}^2 > 0, \quad \sum_{j=1}^{k_n} \sigma_{nj}^2 = 1.$$

Let $S_n = X_{n1} + \dots + X_{nk_n}$, $F_{nj}(x) = P(X_{nj} \leq x)$.

Lemma 3 (*Lindeberg-Feller Theorem*) *The following two statements are equivalent:*

$$S_n \longmapsto N(0, 1) \text{ in distribution,}$$

$$\max_{1 \leq k \leq k_n} EX_{nk}^2 \longmapsto 0 \text{ as } n \mapsto \infty,$$

and

$$(L) \text{ for every } \epsilon > 0; \quad L_n(\epsilon) \equiv \sum_{j=1}^{k_n} \int_{[|x| > \epsilon]} x^2 dF_{nj}(x) \longmapsto 0 \text{ as } n \mapsto \infty.$$

Theorem 4 *Under the notations given above, we have for almost all realizations of the Markov chain $\{X_n; n \geq 0\}$,*

$$\sqrt{N_n}(\tilde{P}_n - \hat{P}_n) \mapsto N(0, \Sigma_P) \text{ in distribution}$$

as $n \mapsto \infty$ and $N_n \mapsto \infty$, where Σ_P is the same variance-covariance matrix as the one in Theorem 3.

Proof. Let

$$\begin{aligned} m_i^{(n)} &\equiv \sum_{r=0}^{N_n} I(X_{nr} = i), \\ m_{ij}^{(n)} &\equiv \sum_{r=0}^{N_n-1} I(X_{nr} = i, X_{n(r+1)} = j). \end{aligned}$$

Then

$$\sqrt{N_n}(\tilde{P}_n - \hat{P}_n) = (\sqrt{N_n}(\frac{m_{ij}^{(n)}}{m_i^{(n)}} - p_{nij}))_{i,j=1,2,\dots,k},$$

where p_{nij} is the ij entry of \hat{P}_n .

By the Cramér-Wold device, Theorem 4 would be proved if we show that for any real numbers l_{ij} ,

$$\sum_{i,j=1}^k l_{ij} \sqrt{N_n} \left(\frac{m_{ij}^{(n)}}{m_i^{(n)}} - p_{nij} \right) \quad (2.8)$$

converges to the normal distribution. In what follows, we shall show only that for each i, j

$$\sqrt{N_n} \left(\frac{m_{ij}^{(n)}}{m_i^{(n)}} - p_{nij} \right) \quad (2.9)$$

converges to the normal distribution. Our proof can be modified easily to establish (2.8).

Now, for each fixed n , given $\underline{X} = (X_0, X_1, \dots, X_n)$, let $\{X_{nr}\}$, $r = 1, \dots, N_n$ be a realization of a Markov chain with transition probability \hat{P}_n . We can view this as having been generated in the following fashion:

Consider an independent collection of random variables $\{W_{it}^{(n)}\}$, where $i = 1, \dots, k$, $t = 1, 2, \dots$ such that

$$P\{W_{it}^{(n)} = j\} = p_{nij}, \quad j = 1, \dots, k.$$

Now, let

$$\begin{aligned} X_0^{(n)} &= 1, \\ X_{i+1}^{(n)} &= W_{X_i^{(n)} m}^{(n)} \quad \text{for } i \geq 0, \end{aligned}$$

where $m - 1$ is the number of l with $1 \leq l \leq N_n$ such that $X_l^{(n)} = X_i^{(n)}$.

Now, for fixed i and j , we define that

$$C_t^{(n)}(j) \equiv \begin{cases} 1, & \text{if } W_{it}^{(n)} = j, \quad t = 1, \dots, m_i^{(n)}; \\ 0, & \text{otherwise,} \end{cases} \quad (2.10)$$

where $m_i^{(n)}$ is the number of visits to i by the n^{th} chain.

By the definition given above, we have that $P(C_t^{(n)}(j) = 1) = p_{nij}$ and $P(C_t^{(n)}(j) = 0) = 1 - p_{nij}$. Let $\hat{\Pi}_n = (\pi_{n1}, \pi_{n2}, \dots, \pi_{nk})$ be the limiting distribution of \hat{P}_n . For fixed i such that $\pi_{ni} > 0$, we define

$$d_t^{(n)}(j) \equiv \frac{C_t^{(n)}(j) - p_{nij}}{\sqrt{[N_n \pi_{ni}] p_{nij} (1 - p_{nij})}}, \quad t = 1, 2, \dots, [N_n \pi_{ni}], \quad n = 1, 2, \dots.$$

Then we have a double array of random variables $\{d_t^{(n)}(j)\}$, $t = 1, 2, \dots$ with $Ed_t^{(n)} = 0$ and $\sigma_{nt}^2(j) \equiv Var(d_t^{(n)}(j)) = 1/[N_n\pi_{ni}]$.

First, we want to show that conditioned on $\mathbf{X} = (X_0, X_1, \dots, X_n)$,

$$S_{[N_n\pi_{ni}]}^{(n)}(j) \equiv \sum_{t=1}^{[N_n\pi_{ni}]} d_t^{(n)}(j) \mapsto N(0, 1) \text{ in distribution.}$$

In order to get this result, we need to check the Lindeberg's condition, that is, we want to show

$$\text{For any given } \epsilon > 0, \quad L_n(\epsilon) \equiv \sum_{t=1}^{[N_n\pi_{ni}]} \int_{[|x|>\epsilon]} x^2 dF_{nt}(x) \mapsto 0 \text{ as } n \mapsto \infty,$$

where $F_{nt}(x)$ is the distribution function of $d_t^{(n)}(j)$.

In fact, we have the following inequality:

$$\begin{aligned} |d_t^{(n)}(j)| &= \frac{|C_t^{(n)}(j) - p_{nij}|}{\sqrt{[N_n\pi_{nj}]p_{nij}(1-p_{nij})}} \\ &\leq \frac{2}{\sqrt{[N_n\pi_{nj}]p_{nij}(1-p_{nij})}}. \end{aligned}$$

Since $\hat{P}_n \mapsto P$ with probability 1 as $n \mapsto \infty$, and since P is a finite irreducible matrix, we have $\hat{\Pi}_n \mapsto \Pi$ with probability 1, where $\Pi = \Pi P$. Thus, as $n \mapsto \infty$, we have $\pi_{nj}p_{nij}(1-p_{nij}) \mapsto \pi_j p_{ij}(1-p_{ij}) > 0$. Now, since $N_n \mapsto \infty$, we have for any given $\epsilon > 0$, with probability 1, there exists $n_0 > 0$ such that for $n > n_0$, $\sup_t d_t^{(n)}(j) \leq \epsilon$.

Therefore

$$\int_{[|x|>\epsilon]} x^2 dF_{nt}(x) = 0 \text{ for all } t = 1, \dots, [N_n\pi_{ni}] \text{ with } n > n_0,$$

this implies that

$$L_n(\epsilon) \equiv \sum_{t=1}^{[N_n \pi_{ni}]} \int_{[|x| > \epsilon]} x^2 dF_{nt}(x) = 0 \text{ for all } n > n_0.$$

Since $p_{nij} \mapsto p_{ij}$ in probability, we have, by Lemma 3, for each i, j

$$\frac{\sum_{t=1}^{[N_n \pi_{ni}]} (C_t^{(n)}(j) - p_{nij})}{\sqrt{[N_n \pi_{ni}] p_{ij} (1 - p_{ij})}} \mapsto N(0, 1) \text{ in distribution.}$$

Next, we let

$$\dot{m}_{ij}^{(n)} \equiv \sum_{t=1}^{[N_n \pi_{ni}]} I(X_{t-1} = i, X_t = j).$$

Note that

$$\begin{aligned} & \frac{m_{ij}^{(n)} - m_i^{(n)} p_{nij}}{\sqrt{m_i^{(n)}}} \\ &= \frac{m_{ij}^{(n)} - \dot{m}_{ij}^{(n)} + \dot{m}_{ij}^{(n)} - [N_n \pi_{ni}] p_{nij} + [N_n \pi_{ni}] p_{nij} - m_i^{(n)} p_{nij}}{\sqrt{m_i^{(n)}}} \\ &= \frac{\sqrt{[N_n \pi_{ni}]}}{\sqrt{m_i^{(n)}}} \left(\frac{\dot{m}_{ij}^{(n)} - [N_n \pi_{ni}] p_{nij}}{\sqrt{[N_n \pi_{ni}]}} \right) \\ &+ \frac{\sqrt{N_n}}{\sqrt{m_i^{(n)}}} \frac{(m_{ij}^{(n)} - m_i^{(n)} p_{nij}) - (\dot{m}_{ij}^{(n)} - [N_n \pi_{ni}] p_{nij})}{\sqrt{N_n}}. \end{aligned}$$

Since

$$\frac{\sum_{t=1}^{[N_n \pi_{ni}]} (C_t^{(n)}(j) - p_{nij})}{\sqrt{[N_n \pi_{ni}]}}$$

has the same distribution as

$$\frac{\dot{m}_{ij}^{(n)} - [N_n \pi_{ni}] p_{nij}}{\sqrt{[N_n \pi_{ni}]}}.$$

Hence, we have

$$\frac{m_{ij}^{(n)} - [N_n \pi_{ni}] p_{nij}}{\sqrt{[N_n \pi_{ni}]}} \mapsto N(0, p_{ij}(1 - p_{ij})) \text{ in distribution.}$$

Next, we verify that for almost all realizations of the process, we have

$$\begin{aligned} a) \quad & \frac{m_i^{(n)}}{N_n} - \pi_{ni} \mapsto 0 \text{ in probability.} \\ b) \quad & \eta_{nij} \equiv \frac{m_{ij}^{(n)} - m_i^{(n)} p_{nij}}{\sqrt{N_n}} - \frac{m_{ij}^{(n)} - [N_n \pi_{ni}^{(n)}] p_{nij}}{\sqrt{N_n}} \mapsto 0 \text{ in probability.} \end{aligned}$$

We prove a) and b) separately, and prove a) first.

In both these paragraphs, P and E will denote the conditional probability and conditional expectation given the observation \mathbf{X} .

a) For each fixed n , for all α , we have $|p_{nij}^{(\alpha)} - \pi_{nj}| \leq C \rho_n^\alpha$ where $C > 0$, is an universal constant and $\rho_n = 1 - \min_{i,j} p_{nij} < 1$. By recurrence and finiteness of the state space, ρ_n converges to $\rho \equiv 1 - \min_{i,j} p_{ij} < 1$. Hence, we can find a number $\rho < 1$, and $n_0(\omega)$ such that for $n \geq n_0(\omega)$,

$$|p_{nij}^{(\alpha)} - \pi_{nj}| \leq C \rho^\alpha \quad \text{for all } \alpha.$$

Now, for each $\epsilon > 0$, we have

$$P\left\{ \left| \frac{m_i^{(n)}}{N_n} - \pi_{ni} \right| > \epsilon \right\} \leq \frac{E\left[\left(\frac{m_i^{(n)}}{N_n} - \pi_{ni}\right)^2\right]}{\epsilon^2},$$

and

$$E\left[\left(\frac{m_i^{(n)}}{N_n} - \pi_{ni}\right)^2\right]$$

$$\begin{aligned}
&= E\left[\frac{1}{N_n^2} \left(\sum_{k=1}^{N_n} (I_i(X_{nk}) - \pi_{ni}) \right)^2\right] \\
&= \frac{1}{N_n^2} E\left(\sum_{k=1}^{N_n} \sum_{l=1}^{N_n} m_{ni}^{(k,l)} \right),
\end{aligned}$$

where

$$m_{ni}^{(k,l)} = I_i(X_{nk})I_i(X_{nl}) - \pi_{ni}I_i(X_{nl}) - \pi_{ni}I_i(X_{nk}) + \pi_{ni}^2.$$

Hence

$$Em_{ni}^{(k,l)} = p_{n1i}^{(s)} p_{nii}^{(t)} - \pi_{ni} p_{n1i}^{(l)} - \pi_{ni} p_{n1i}^{(k)} + \pi_{ni}^2,$$

where $s = \min(k, l)$, $t = |k - l|$.

We have that $p_{nii}^{(\alpha)} = \pi_{ni} + \epsilon_{ni}^{(\alpha)}$ where $|\epsilon_{ni}^{(\alpha)}| \leq C\rho^\alpha$.

Therefore

$$\begin{aligned}
&Em_{ni}^{(k,l)} \\
&= (\pi_{ni} + \epsilon_{ni}^{(s)})(\pi_{ni} + \epsilon_{ni}^{(t)}) - \pi_{ni}(\pi_{ni} + \epsilon_{ni}^{(l)}) - \pi_{ni}(\pi_{ni} + \epsilon_{ni}^{(k)}) + \pi_{ni}^2 \\
&= \pi_{ni}^2 + \pi_{ni}\epsilon_{ni}^{(s)} + \pi_{ni}\epsilon_{ni}^{(t)} + \epsilon_{ni}^{(s)}\epsilon_{ni}^{(t)} - \pi_{ni}^2 - \pi_{ni}\epsilon_{ni}^{(l)} - \pi_{ni}^2 - \pi_{ni}\epsilon_{ni}^{(k)} + \pi_{ni}^2 \\
&= \pi_{ni}\epsilon_{ni}^{(s)} + \pi_{ni}\epsilon_{ni}^{(t)} + \epsilon_{ni}^{(s)}\epsilon_{ni}^{(t)} - \pi_{ni}\epsilon_{ni}^{(l)} - \pi_{ni}\epsilon_{ni}^{(k)}.
\end{aligned}$$

This implies that, for $n \geq n_0(\omega)$,

$$\begin{aligned}
&|Em_{ni}^{(k,l)}| \\
&\leq \pi_{ni} [|\epsilon_{ni}^{(s)}| + |\epsilon_{ni}^{(t)}| + C_1 |\epsilon_{ni}^{(s)}\epsilon_{ni}^{(t)}| + |\epsilon_{ni}^{(l)}| + |\epsilon_{ni}^{(k)}|] \quad (C_1 = 1/\pi_{ni}) \\
&\leq C_2 [\rho^s + \rho^t + C_1 C \rho^s \rho^t + \rho^k + \rho^l] \\
&\leq C [\rho^s + \rho^t + \rho^k + \rho^l] \text{ for some constant } C > 0.
\end{aligned}$$

Therefore

$$\begin{aligned}
& E[(\frac{m_i^{(n)}}{N_n} - \pi_{ni})^2] \\
& \leq \frac{1}{N_n^2} \sum_{k=1}^{N_n} \sum_{l=1}^{N_n} C[\rho^s + \rho^t + \rho^k + \rho^l] \quad (\text{here we assume } l > k) \\
& \leq \frac{C}{N_n^2} \sum_{k=1}^{N_n} \sum_{l=1}^{N_n} [2\rho^k + \rho^{l-k} + \rho^l] \\
& \leq \frac{C}{N_n^2} [3N_n \sum_{k=1}^{\infty} \rho^k + \sum_{l=1}^{N_n} \sum_{k=1}^l \rho^{l-k}] \\
& \leq \frac{C}{N_n^2} [3N_n \frac{1}{1-\rho} + N_n \frac{1}{1-\rho}] \\
& \leq \frac{4N_n C}{N_n^2(1-\rho)} \\
& = \frac{4C}{N_n(1-\rho)}.
\end{aligned}$$

This implies that

$$P\{|\frac{m_i^{(n)}}{N_n} - \pi_{ni}| > \epsilon\} \leq \frac{4C}{\epsilon^2 N_n(1-\rho)} \longrightarrow 0 \text{ as } n \mapsto \infty.$$

b) Let

$$S_m \equiv \sum_{t=1}^m (C_t^{(n)}(j) - p_{nij}), \quad \text{where } C_t^{(n)}(j) \text{ is as in (2.10).}$$

By a), for any given $\epsilon > 0$, there exists a positive integer $n_0(\omega)$, such that

$$P\{|\frac{m_i^{(n)}}{N_n} - \pi_{ni}| > \epsilon^3 N_n\} < \epsilon \text{ for all } n > n_0(\omega).$$

Consider

$$P\{|\eta_{nij}| > \epsilon\}$$

$$\begin{aligned}
&= P\{|S_{m_i^{(n)}} - S_{[N_n \pi_{ni}]}| > \sqrt{N_n \epsilon}\} \\
&= P\{|m_i^{(n)} - [N_n \pi_{ni}]| > \epsilon^3 N_n, |S_{m_i^{(n)}} - S_{[N_n \pi_{ni}]}| > \sqrt{N_n \epsilon}\} \\
&\quad + P\{|m_i^{(n)} - [N_n \pi_{ni}]| \leq \epsilon^3 N_n, |S_{m_i^{(n)}} - S_{[N_n \pi_{ni}]}| > \sqrt{N_n \epsilon}\} \\
&\leq P\{|m_i^{(n)} - [N_n \pi_{ni}]| > \epsilon^3 N_n\} \\
&\quad + P\{\max_{|m - [N_n \pi_{ni}]| \leq \epsilon^3 N_n} |S_m - S_{[N_n \pi_{ni}]}| > \sqrt{N_n \epsilon}\} \\
&\leq \epsilon + 2P\{\max_{1 \leq m \leq \epsilon^3 N_n} |S_m| > \sqrt{N_n \epsilon}\} \\
&\leq \epsilon + 2 \frac{2}{\epsilon^2 N_n} \text{Var}(S_{[N_n \epsilon^3] + 1}) \quad (\text{by Kolmogorov's inequality}) \\
&= \epsilon + \frac{4}{\epsilon^2 N_n} ([N_n \epsilon^3] + 1) \sigma^2 \\
&\leq C \epsilon \quad (\text{where } C > 0 \text{ is a constant}).
\end{aligned}$$

This proves b).

By definition of $m_{ij}^{(n)}$ which is the sum of the i.i.d. sequence $C_t^{(n)}(j)$ for each n , we have for each i, j

$$\frac{m_{ij}^{(n)} - [N_n \pi_{ni}] p_{nij}}{\sqrt{[N_n \pi_{ni}]}} \mapsto N(0, \sigma^2) \text{ in distribution.}$$

Hence, by a) and b) we have that

$$\frac{m_{ij}^{(n)} - m_i^{(n)} p_{nij}}{\sqrt{m_i^{(n)}}} \mapsto N(0, p_{ij}(1 - p_{ij})) \text{ in distribution.}$$

Finally,

$$\sqrt{N_n}(\tilde{p}_{nij} - p_{nij}) = \sqrt{N_n} \left(\frac{m_{ij}^{(n)}}{m_i^{(n)}} - p_{nij} \right)$$

$$= \frac{\sqrt{N_n} \frac{m_{ij}^{(n)} - m_i^{(n)} p_{nij}}{\sqrt{m_i^{(n)}}}}{\sqrt{m_i^{(n)}}}$$

and this converges in distribution to $N(0, \frac{p_{ij}(1-p_{ij})}{\pi_i})$. \square

2.4 Bootstrapping the Hitting Time of a Markov Chain

Consider the same situation as in Section 2.2. Let T_k be the first hitting time of a given state k . That is, we let

$$T_k = \begin{cases} \inf\{n \geq 0; X_n = k\}; \\ \infty, & \text{if no such } n \text{ exist.} \end{cases}$$

Let $Pr(t; P) \equiv P(T_k \leq t \mid X_0 = 1; P)$ denote the probability that $T_k \leq t$ for $t \in \{1, 2, 3, \dots\}$, where P is the transition probability matrix of a Markov chain $X = \{X_n; n \geq 0\}$ with initial state $X_0 = 1$.

Definition 3 For any stochastic matrix P , let $A \equiv A(P)$ be the stochastic matrix which is the same as P except that the k^{th} row is replaced by $(0, \dots, 0, 1, 0, \dots, 0)$ with 1 in the k^{th} position.

Note that

$$Pr(t; P) = (A^t)_{1,k} \tag{*}$$

The bootstrap estimate of the distribution $Pr(t; P)$ of the hitting time T_k is $Pr(t; \hat{P}_n)$. From (*) and $\hat{P}_n \mapsto P$ in probability, we have

$$Pr(t; \hat{P}_n) - Pr(t; P) \mapsto 0 \text{ in probability as } n \mapsto \infty.$$

By Theorem 5 below we estimate the distribution of

$$G_n(t; P) \equiv \sqrt{n}(Pr(t; \hat{P}_n) - Pr(t; P)) \quad t \geq 1. \quad (2.11)$$

The bootstrap approximation to the distribution in (2.11) is the distribution of

$$G_n(t; \hat{P}_n) \equiv \sqrt{N_n}(Pr(t; \tilde{P}_n) - Pr(t; \hat{P}_n)). \quad (2.12)$$

The problem here is to verify that these two distributions are asymptotically close, which leads to Theorem 6.

Theorem 5 *Let $\hat{A}_n \equiv A(\hat{P}_n)$. Then for all $t = 1, 2, 3, \dots$, we have*

$$\sqrt{n}(\hat{A}_n^t - A^t) \mapsto N(0, Z_P^t) \text{ in distribution,}$$

where the variance-covariance matrix Z_P^t is continuous as a function of P with respect to the supremum norm on the class of $k \times k$ stochastic matrices.

Proof. First, we have the following identity which can be proved by mathematical induction,

$$\sqrt{n}(\hat{A}_n^t - A^t) = \sum_{q=0}^{t-1} \hat{A}_n^{t-1-q} \sqrt{n}(\hat{A}_n - A) A^q.$$

By Theorem 3, we have

$$\sqrt{n}(\hat{P}_n - P) \mapsto N(0, \Sigma_P) \text{ in distribution,}$$

this implies that

$$\sqrt{n}(\hat{A}_n - A) \mapsto N(0, \Sigma_A) \text{ in distribution,}$$

where Σ_A is the variance covariance matrix. Since \hat{A}_n converges to A in probability, this implies

$$\hat{A}_n^{t-1-q} - A^{t-1-q} \mapsto 0 \text{ in probability.}$$

Therefore, by Slutsky's theorem, we have

$$\sqrt{n}(\hat{A}_n^t - A^t) \mapsto \sum_{q=0}^{t-1} A^{t-1-q} U A^q \text{ in distribution,}$$

where $U \sim N(0, \Sigma_A)$. We now define Z_P^t to be the variance-covariance of the right side of the above equation. Note that for $t = 1, 2$ we have

$$\begin{aligned} Z_P^1 &= \Sigma_A, \\ Z_P^2 &= \text{Var}(AU + UA). \quad \square \end{aligned}$$

Theorem 6 *Let $\tilde{A}_n \equiv A(\tilde{P}_n)$ where \tilde{P}_n is defined in (2.6). Then, for almost all realizations of the process, and for all $t = 1, 2, 3, \dots$, we have*

$$\sqrt{N_n}(\tilde{A}_n^t - \hat{A}_n^t) \mapsto N(0, Z_P^t) \text{ in distribution,}$$

where Z_P^t is the same variance-covariance matrix as the one in Theorem 5.

Proof. From Theorem 4, we have that for almost all realizations of the process,

$$\sqrt{N_n}(\tilde{P}_n - \hat{P}_n) \mapsto N(0, \Sigma_P) \text{ in distribution.}$$

The rest of the proof is the same as the one in Theorem 5. \square

Adapting the same type of analysis, one can estimate

$m(P) \equiv E(T_k \mid X_0 = 1; P)$ the expected value of the hitting time T_k by

$m(\hat{P}_n) \equiv E(T_k \mid X_0 = 1; \hat{P}_n)$. It will be shown below that $m(\hat{P}_n)$ is a consistent estimator of $m(P)$, and further

$$Q_n(t; P) \equiv \sqrt{n}(m(\hat{P}_n) - m(P)) \text{ is asymptotically normal.} \quad (2.13)$$

The bootstrap approximation to the distribution of the left side of (2.13) is the distribution of

$$Q_n(t; \hat{P}_n) \equiv \sqrt{N_n}(m(\tilde{P}_n) - m(\hat{P}_n)). \quad (2.14)$$

The problem here is to justify that these two distributions are asymptotically close. In order to do this, we need the following lemmas which are important by themselves.

Lemma 4 (*δ -method*) Suppose $\{X_n\}$ is a sequence of random k -vectors such that $\sqrt{n}(X_n - \underline{c})$ converges in distribution to the k -variate normal distribution, with mean $\underline{0}$ and variance-covariance matrix Σ . Let f be a real-valued function from R^k to R such that f has continuous first order partial derivatives at \underline{c} . Then,

$$\sqrt{n}(f(X_n) - f(\underline{c})) \mapsto N(0, \sigma^2) \text{ in distribution,}$$

where $\sigma^2 = (f'(\underline{c}))' \Sigma f'(\underline{c})$ and $f'(\underline{c})$ is the column vector with entries $\frac{df(\underline{c})}{dx_j}$, for $j = 1, 2, \dots, k$.

Lemma 5 Let P be a stochastic $k \times k$ matrix and $\{X_n; n \geq 0\}$ be a Markov chain with transition probability matrix P . Let

$$m_{ijk}(P) \equiv E(\text{visits to } j \text{ before to } k \mid X_0 = i; P).$$

Then

$$m_{ijk}(P) = \sum_{r=0}^{\infty} (P_k^r)_{ij},$$

where P_k is the matrix given by P in which the k^{th} column is replaced by $\underline{0}$. Note that $m_{ijk}(P)$ is continuously differentiable with respect to the entries of P at all P such that $\inf_{i,j} p_{ij} > 0$.

Remark: Note that $m(P) = \sum_{j \in S} m_{1j} k(P)$ which is also continuously differentiable with respect to the entries of P at all P such that $\inf_{i,j} p_{ij} > 0$. By the δ -method, Theorem 3 and Theorem 4, we have both

$$\begin{aligned} \sqrt{n}(m(\hat{P}_n) - m(P)) &\longmapsto N(0, \sigma_P^2) \text{ in distribution,} \\ \text{and } \sqrt{N_n}(m(\tilde{P}_n) - m(\hat{P}_n)) &\longmapsto N(0, \sigma_P^2) \text{ in distribution.} \end{aligned}$$

where σ_P^2 is the variance of $m(P)$, which is continuous as a function of P with respect to the supremum norm on the class of $k \times k$ stochastic matrices.

From the above argument, we prove that the bootstrap method works well for estimating the distribution of

$$G_n(t; P) \equiv \sqrt{n}(Pr(t; \hat{P}_n) - Pr(t; P)),$$

and the distribution of

$$Q_n(t; P) \equiv \sqrt{n}(m(\hat{P}_n) - m(P)).$$

3 BOOTSTRAPPING AN INFINITE STATE MARKOV CHAIN

3.1 Introduction

In Chapter 2, we discussed the bootstrap method for estimating the distribution of $\sqrt{n}(\hat{P}_n - P)$ for an ergodic Markov chain with transition probability matrix P and finite state space. The generalization from finite state space to infinite state space will be investigated in this chapter.

The bootstrap algorithm for estimating the distribution of $\sqrt{n}(\hat{P}_n - P)$ depends on both the estimator \hat{P}_n of P as well as the bootstrap resample size. In this chapter, we propose three different estimators \hat{P}_n of P and the corresponding bootstrap methods.

Estimator I: The same algorithm as described in Section 2.3.

Before we introduce the 2nd and 3rd estimators, let us consider the idea of a regeneration process. The existence of a recurrent state Δ which is visited infinitely often (i.o.) for a recurrent Markov chain is well known. A famous approach to its limit theory is via the embedded renewal process of returns to Δ . This is the so-called regeneration method. For a fixed state Δ , by the strong Markov property, the cycles $\{X_j; j = T_{\Delta}^{(n)}, \dots, T_{\Delta}^{(n+1)} - 1\}$ are i.i.d. for $n = 1, 2, \dots$, where $T_{\Delta}^{(n)}$ is the time of the n^{th} return to Δ .

Estimator II: Fix an integer k and observe the chain up to the random time

$n = T_{\Delta}^{k+1}$. Let

$$\{X_0, X_1, \dots, X_n\}$$

be a realization of the process. Note that in this situation, $X_n = \Delta$. Fix i, j , let $\eta_{\alpha} \equiv \{X_j; j = T_{\Delta}^{\alpha}, \dots, T_{\Delta}^{\alpha+1} - 1\}$ denote the α^{th} cycle, $g(\eta_{\alpha})$ indicate the number of visits to state i during the cycle η_{α} , and $h(\eta_{\alpha})$ indicate the number of ij transitions during the cycle η_{α} . Now, define

$$\hat{\pi}_k(i) \equiv \frac{\sum_{\alpha=1}^k g(\eta_{\alpha})}{\sum_{\alpha=1}^k T_{\alpha}}, \quad \hat{p}_k(i, j) \equiv \frac{\sum_{\alpha=1}^k h(\eta_{\alpha})}{\sum_{\alpha=1}^k g(\eta_{\alpha})}$$

be the estimators of Π and P , where T_{α} is the length of η_{α} .

The bootstrap algorithm is as follows:

1) The original sample can be decomposed in the following fashion:

$$\{\eta_0, \eta_1, \eta_2, \dots, \eta_k\}, \text{ where } \eta_0 \equiv \{X_0, X_1, \dots, X_{T_{\Delta}^{(1)} - 1}\}.$$

Let \hat{F}_k denote the uniform probability measure on the cycles $\{\eta_{\alpha}; \alpha = 1, 2, \dots, k\}$.

2) With the original sample fixed, draw a “bootstrap sample” of size k according to \hat{F}_k . Denote this sample by $\eta_1^*, \eta_2^*, \dots, \eta_k^*$. Then, the bootstrap estimators of $\hat{\pi}_k(i)$, $\hat{p}_k(i, j)$ can be defined as follows:

$$\tilde{\pi}_k(i) \equiv \frac{\sum_{\alpha=1}^k g(\eta_{\alpha}^*)}{\sum_{\alpha=1}^k T_{\alpha}^*}, \quad \tilde{p}_k(i, j) \equiv \frac{\sum_{\alpha=1}^k h(\eta_{\alpha}^*)}{\sum_{\alpha=1}^k g(\eta_{\alpha}^*)},$$

where T_{α}^* is the length of η_{α}^* .

3) Approximate the distribution of $R(\underline{X}, F) \equiv \sqrt{k}(\hat{p}_k(i, j) - p_{ij})$ by the conditional distribution of $R(\underline{X}^*, \hat{F}_k) \equiv \sqrt{k}(\tilde{p}_k(i, j) - \hat{p}_k(i, j))$ given \underline{x} .

The details will be discussed in Section 3.3.

Estimator III: Fix the original sample size n . Let k be the (random) number of full cycles included in the observation $\{X_0, X_1, \dots, X_n\}$. The bootstrap method here is to use these k cycles. We will investigate this case in Section 3.4.

3.2 Estimator I

Let $X \equiv \{X_n; n \geq 0\}$ be a homogeneous ergodic Markov chain with transition probability matrix $P = (p_{ij})$ and countable infinite state space S . Then there exists an invariant probability measure $\Pi = (\pi_1, \pi_2, \dots)$ such that $\pi_j > 0$, $\sum_{j=1}^{\infty} \pi_j = 1$, $p_{ij}^{(n)} \mapsto \pi_j$, as $n \rightarrow \infty$ for all $i \in S$ and $\pi_j = \sum_{\alpha} \pi_{\alpha} p_{\alpha j}$, for all $j = 1, 2, \dots$.

Suppose $\mathbf{x} \equiv \{x_0, x_1, \dots, x_n\}$ is a realization of the process observed up to time n . We estimate P by its maximum likelihood estimator $\hat{P}_n \equiv (\hat{p}_n(i, j))$, where

$$\hat{p}_n(i, j) = \begin{cases} n_{ij}/n_i, & \text{if } n_i > 0; \\ \delta_{ij}, & \text{otherwise.} \end{cases} \quad (3.1)$$

and estimate Π by $\hat{\Pi}_n \equiv (\hat{\pi}_n(i))$, where

$$\hat{\pi}_n(i) = \frac{n_i}{n}, \quad (3.2)$$

n_{ij} = observed number of ij transitions in $\{x_0, \dots, x_n\}$,

n_i = observed number of visits to state i in $\{x_0, \dots, x_n\}$.

It is clear that a finite number of observations will not provide an estimator for all states of the transition probabilities. It is well known that for an irreducible positive recurrent Markov chain, the range space R_n of $\{X_0, X_1, \dots, X_n\}$ is asymptotically small compared to the number of observations n . In mathematical terms, it can be described as follows:

Theorem 7 *Let $\{X_n; n = 0, 1, \dots\}$ be an irreducible recurrent Markov chain on the nonnegative integers. Let $R_n \equiv$ cardinality of $\{X_0, X_1, \dots, X_n\}$ be the range sequence. Let*

$$T_\Delta \equiv \inf\{n : n \geq 1, X_n = \Delta\},$$

be the first hitting time of the state Δ , and λ be any initial distribution such that $E_\lambda(R_{T_\Delta}) < \infty$, then

$$\frac{R_n}{n} \longmapsto 0 \quad w.p.1.$$

See Athreya [3] for the proof.

The consistency and the asymptotic normality of the estimators $\hat{\Pi}_n$ and \hat{P}_n are well known, see Derman [16] and Nummelin [33] for the details. We state the following two theorems here for reference.

Theorem 8 *Let X be a homogeneous irreducible positive recurrent Markov chain. Then, for each i*

$$\hat{\pi}_n(i) \longmapsto \pi_i \quad \text{with probability 1 as } n \mapsto \infty.$$

Theorem 9 *Let X be a homogeneous irreducible positive recurrent Markov chain with $E_\Delta T_\Delta^2 < \infty$. Then*

$$\sqrt{n}(\hat{P}_n - P) \longmapsto N(0, \Sigma_P) \quad \text{in distribution,}$$

where Σ_P , the variance-covariance matrix, is continuous as a function of P with respect to the supremum norms on the class of stochastic matrices and on the class of variance covariance matrices.

Here convergence in distribution means that for any finite set A of pairs (i, j) ,

$$\{\sqrt{n}(\hat{p}_n(i, j) - p_{ij}); (i, j) \in A\} \mapsto N(0, (\Sigma_P)_A),$$

where $(\Sigma_P)_A$ is a block diagonal matrix involving the states in A .

Let \tilde{P}_n be the bootstrap estimator defined as before. The problem here is to verify that $\sqrt{n}(\hat{P}_n - P)$ and $\sqrt{N_n}(\tilde{P}_n - \hat{P}_n)$ have the same asymptotic behavior. This is done as follows:

With the same set up as above, we want to verify that

$$\sqrt{N_n}(\tilde{P}_n - \hat{P}_n) \mapsto N(0, \Sigma_P) \text{ in distribution}$$

in a suitable sense, that is, either for almost all realizations of the original chain or in probability.

If we follow the proof of Theorem 4, the only step that needs a change is the one to show that for all $\epsilon > 0$

$$P(|\frac{m_i^{(n)}}{N_n} - \pi_{ni}| > \epsilon \mid \underline{X}) \mapsto 0 \quad w.p.1. \text{ (or in probability)}. \quad (3.3)$$

In the finite Markov chain case (Theorem 4), we exploited the geometric ergodicity of $(p_{ij}^{(n)} - \pi_j)$ and the finiteness of the state space in showing that

$$E(|\frac{m_i^{(n)}}{N_n} - \pi_{ni}|^2 \mid \underline{X}) \mapsto 0 \quad w.p.1. \quad \text{as } n, N_n \mapsto \infty.$$

Unfortunately, this does not carry over to the infinite state space case. We do still have that

$$\begin{aligned} & P(|\frac{m_i^{(n)}}{N_n} - \pi_{ni}| > \epsilon \mid \underline{X}) \\ & \leq \frac{1}{\epsilon^2} (\frac{1}{4N_n} + \frac{1}{N_n^2} \sum_{k=1}^{N_n} \sum_{s=1}^{N_n-k} p_{n1j}^{(k)} (p_{nij}^{(s)} - p_{n1j}^{(s+k)})). \end{aligned}$$

What we need to show is that

$$\Delta_n(\omega) \equiv \frac{1}{N_n^2} \sum_{k=1}^{N_n} \sum_{s=1}^{N_n-k} p_{n1j}^{(k)} (p_{nij}^{(s)} - p_{n1j}^{(k+s)}) \mapsto 0 \quad (3.4)$$

w.p.1. (or in probability).

The open question here is as follows:

$$\text{Is it true that } \Delta_n(\omega) \mapsto 0 \text{ w.p.1. (or in probability)?}$$

A sufficient condition is that for all $\epsilon > 0$

$$\sup_{s \geq \epsilon N_n} |p_{n1j}^{(s)} - \pi_{nj}| \mapsto 0 \quad \text{w.p.1. (or in probability).}$$

At this moment, we do not know a nice sufficient condition to ensure (3.4). In practice, it is possible to choose N_n depending on \hat{P}_n to ensure that (3.4) holds.

It is interesting to note that Theorem 9 needs $E_{\Delta} T_{\Delta}^2 < \infty$. We do not know whether $E_{\Delta} T_{\Delta}^2 < \infty$ implies (3.4).

For an ergodic Markov chain with stochastic matrix P , let i be any state and fix n . Under the same notations given above, for any given $\epsilon > 0$, let

$$\phi(\epsilon, n, P) \equiv P(|\hat{\pi}_n(i) - \pi_i| > \epsilon).$$

Then, (3.4) can be reformulated as follows:

$$\phi(\epsilon, n, N_n, \hat{P}_n) \mapsto 0 \quad \text{w.p.1. (or in probability).} \quad (3.5)$$

An interesting open question for the convergence in (3.5) is to determine conditions on $\alpha(P)$, where $\alpha(P)$ is the ergodic coefficient which can be defined as follows:

Definition 4 Let P be a stochastic matrix. The ergodic coefficient of P , denoted by $\alpha(P)$, is defined by

$$\alpha(P) = 1 - \sup_{i,k} \sum_{j=1}^{\infty} [p_{ij} - p_{kj}]^+,$$

where $[p_{ij} - p_{kj}]^+ = \max(0, p_{ij} - p_{kj})$.

Isaacson and Madsen [29] is a good reference on the concept of the ergodic coefficient.

3.3 Estimator II

Recall that estimator II is based on k full cycles. Our goal is to estimate the distribution of $\sqrt{k}(\hat{\pi}_k(i) - \pi_i)$ and $\sqrt{k}(\hat{p}_k(i, j) - p_{ij})$. We begin by recalling some well known results for the i.i.d. case.

Let X_1, X_2, \dots be i.i.d. random variables with $EX_1 = \mu < \infty$. Given a realization $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and integers n and m , construct $Y_{n,i}$, $i = 1, 2, \dots, m$ as i.i.d. random variables with conditional distribution $P^*(Y_{n,i} = X_j) = 1/n$ for $1 \leq j \leq n$. (P^* denotes conditional distribution given \mathbf{X}). We have the following theorem regarding the strong law for the bootstrap. We state the result without proof, the reader is referred to Athreya [3] for the details.

Theorem 10 If $\liminf mn^{-\beta} > 0$ for some $\beta > 0$ as $m, n \mapsto \infty$ and $E |X_1 - \mu|^\theta < \infty$ for some $\theta \geq 1$ such that $\theta\beta > 1$, then

$$\frac{1}{m} \sum_{i=1}^m Y_{n,i} \xrightarrow{w.p.1.} \mu \text{ as } m, n \mapsto \infty.$$

The strong consistency property of the bootstrap estimator $\tilde{\pi}_k(i)$ is given by the following theorem.

Theorem 11 *With the same notations given above, if there exists $\delta > 0$ such that $E_{\Delta} T_{\Delta}^{1+\delta} < \infty$, then*

$$\tilde{\pi}_k(i) \longmapsto \pi_i \text{ w.p.1. as } k \mapsto \infty.$$

Proof. Recall that

$$\hat{\pi}_k(i) = \frac{\sum_{t=1}^k g(\eta_t)}{\sum_{t=1}^k T_t},$$

where

$$g(\eta_t) = \text{number of visits to } i \text{ during cycle } \eta_t,$$

$$T_t = \text{length of the cycle } \eta_t.$$

The bootstrap version of this is

$$\tilde{\pi}_k(i) \equiv \frac{\sum_{t=1}^k g(\eta_t^*)}{\sum_{t=1}^k T_t^*}.$$

By assumption, we have

$$E_{\Delta} |g(\eta_1)|^{1+\delta} \leq E_{\Delta} T_{\Delta}^{1+\delta} < \infty \quad \text{for some } \delta > 0.$$

By Theorem 10, we have

$$\frac{1}{k} \sum_{t=1}^k g(\eta_t^*) \longmapsto Eg(\eta_1) \text{ with probability 1 as } k \mapsto \infty,$$

and

$$\frac{T_1^* + \cdots + T_k^*}{k} \longmapsto ET_1 \text{ with probability 1 as } k \mapsto \infty.$$

Hence, we have that

$$\tilde{\pi}_k(i) \longmapsto \frac{Eg(\eta_1)}{ET_1} \equiv \pi_i. \quad \square$$

It is well known that

$$\sqrt{k}(\hat{\pi}_k(i) - \pi_i) \text{ is asymptotically normal,}$$

and its bootstrap estimator is the conditional distribution of $\sqrt{k}(\tilde{\pi}_k(i) - \hat{\pi}_k(i))$ given (η_1, \dots, η_k) . The following theorem is an asymptotic one for the classical i.i.d. case, and the reader is referred to Bickel and Freedman [11] and Singh [38] for the details. More recent results are in Hall [27].

Theorem 12 *Suppose X_1, X_2, \dots are independent, identically distributed, and have finite positive variance σ^2 . Along almost all sample sequences X_1, X_2, \dots , given (X_1, \dots, X_n) , as n and m tend to ∞ :*

$$\sqrt{m}(\mu_m^* - \mu_n) \mapsto N(0, \sigma^2) \text{ in distribution,}$$

where $\mu_n(\mu_m^*)$ is the sample mean (bootstrap mean).

The bootstrap method works well to estimate the distribution of $\sqrt{k}(\hat{\pi}_k(i) - \pi_i)$ for an ergodic Markov chain with infinite state space. The statement and proof are as follows:

Theorem 13 *With the notations given above, if $E_\Delta T_\Delta^2 < \infty$, then, for almost all realizations of $\{X_n; n \geq 0\}$, we have*

$$\sqrt{k}(\tilde{\pi}_k(i) - \hat{\pi}_k(i)) \mapsto N(0, \sigma^2) \text{ in distribution as } k \mapsto \infty,$$

where σ^2 is the variance which will be defined in the proof.

Proof. Note that

$$\begin{aligned} \hat{\pi}_k(i) &\equiv \frac{\sum_{t=1}^k g(\eta_t)}{\sum_{t=1}^k T_t}, \\ \tilde{\pi}_k(i) &\equiv \frac{\sum_{t=1}^k g(\eta_t^*)}{\sum_{t=1}^k T_t^*}. \end{aligned}$$

Thus,

$$\begin{aligned}
& \sqrt{k} (\tilde{\pi}_k(i) - \hat{\pi}_k(i)) \\
&= \sqrt{k} \left(\frac{\sum_{t=1}^k g(\eta_t^*)}{\sum_{t=1}^k T_t^*} - \frac{\sum_{t=1}^k g(\eta_t)}{\sum_{t=1}^k T_t} \right) \\
&= \sqrt{k} \left(\frac{1}{\sum_{t=1}^k T_t} \right) \left(\sum_{t=1}^k g(\eta_t^*) \right) \left(\frac{\sum_{t=1}^k T_t}{\sum_{t=1}^k T_t^*} - 1 \right) \\
&\quad + \frac{1}{1/k \sum_{t=1}^k T_t} \frac{1}{\sqrt{k}} \sum_{t=1}^k (g(\eta_t^*) - g(\eta_t)) \\
&= \left(\frac{1/k \sum g(\eta_t^*)}{(1/k \sum T_t)(1/k \sum T_t^*)} - \frac{Eg(\eta_1)}{(ET_1)^2} \right) \left(\frac{1}{\sqrt{k}} \sum (T_t - T_t^*) \right) \\
&\quad + \left(\frac{1}{1/k \sum T_t} - \frac{1}{ET_t} \right) \left(\frac{1}{\sqrt{k}} \sum (g(\eta_t^*) - g(\eta_t)) \right) \\
&\quad + \left[\frac{Eg(\eta_1)}{(ET_1)^2} \frac{1}{\sqrt{k}} \sum (T_t - T_t^*) + \left(\frac{1}{\sqrt{k} ET_t} \sum (g(\eta_t^*) - g(\eta_t)) \right) \right] \\
&= Z_{k1} + Z_{k2} + Z_{k3}, \quad \text{say.}
\end{aligned}$$

By Theorem 11 and strong law, we have

$$\frac{1}{k} \sum_{t=1}^k T_t^* \mapsto ET_1 \quad \text{and} \quad \frac{1}{k} \sum_{t=1}^k g(\eta_t^*) \mapsto Eg(\eta_1).$$

Therefore Z_{k1} and Z_{k2} of the above equation converge to 0 with probability 1 as $k \mapsto \infty$. Let $a \equiv 1/ET_t$ and $b \equiv -Eg(\eta_1)/(ET_1)^2$, then Z_{k3} of the above equation is equal to

$$\begin{aligned}
& \frac{1}{\sqrt{k}} \sum_{t=1}^k (a(g(\eta_t^*) - g(\eta_t)) + b(T_t^* - T_t)) \\
&= \frac{1}{\sqrt{k}} \sum_{t=1}^k ((ag(\eta_t^*) + bT_t^*) - (ag(\eta_t) + bT_t)) \\
&\mapsto N(0, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\sigma_{12}) \text{ in distribution,}
\end{aligned}$$

where $\sigma_1^2 \equiv \text{Var}(T_1)$, $\sigma_2^2 \equiv \text{Var}(g(\eta_1))$ and $\sigma_{12}^2 \equiv \text{Cov}(T_1, g(\eta_1))$. \square

We have the asymptotic normality of $\sqrt{k}(\hat{p}_k(i, j) - p_{ij})$. The asymptotic distribution of $\sqrt{k}(\tilde{p}_k(i, j) - \hat{p}_k(i, j))$ will be investigated in the following theorem.

Theorem 14 *With the notations given above, if $E_\Delta T_\Delta^2 < \infty$, then, for almost all realizations of the process $\{X_n; n \geq 0\}$, we have*

$$\sqrt{k}(\tilde{p}_k(i, j) - \hat{p}_k(i, j)) \mapsto N(0, \sigma^2) \text{ in distribution,}$$

where σ^2 will be defined in the proof.

Proof. We have

$$\begin{aligned} & \sqrt{k} [\tilde{p}_k(i, j) - \hat{p}_k(i, j)] \\ = & \sqrt{k} \left[\frac{\sum h(\eta_t^*)}{\sum g(\eta_t^*)} - \frac{\sum h(\eta_t)}{\sum g(\eta_t)} \right] \\ = & \frac{\sqrt{k}}{\sum g(\eta_t)} \sum h(\eta_t^*) \left[\frac{\sum g(\eta_t)}{\sum g(\eta_t^*)} - 1 \right] + \sqrt{k} \frac{1}{\sum g(\eta_t)} [\sum h(\eta_t^*) - \sum h(\eta_t)] \\ = & \left[\frac{1/k \sum h(\eta_t^*)}{1/k \sum g(\eta_t) 1/k \sum g(\eta_t^*)} - \frac{Eh(\eta_1)}{(Eg(\eta_1))^2} \right] \frac{1}{\sqrt{k}} [\sum g(\eta_t) - g(\eta_t^*)] \\ & + \left[\frac{1}{\sum g(\eta_t)/k} - \frac{1}{Eg(\eta_1)} \right] \frac{1}{\sqrt{k}} [\sum h(\eta_t^*) - \sum h(\eta_t)] \\ & + \left(\frac{1}{Eg(\eta_1)} \frac{1}{\sqrt{k}} [\sum h(\eta_t^*) - \sum h(\eta_t)] - \frac{Eh(\eta_1)}{(Eg(\eta_1))^2} \frac{1}{\sqrt{k}} [\sum g(\eta_t^*) - g(\eta_t)] \right) \\ = & W_{k1} + W_{k2} + W_{k3}, \quad \text{say.} \end{aligned}$$

We have that W_{k1}, W_{k2} converge to 0 with probability 1. Let

$$c \equiv \frac{1}{Eg(\eta_1)}, \quad d \equiv -\frac{Eh(\eta_1)}{Eg(\eta_1)^2}.$$

Then W_{k3} of the above equation is equal to

$$\frac{1}{\sqrt{k}} \sum_{t=1}^k [c(h(\eta_t^*) - h(\eta_t)) + d(g(\eta_t^*) - g(\eta_t))]$$

$$\begin{aligned}
&= \frac{1}{\sqrt{k}} \sum_{t=1}^k [(ch(\eta_t^*) + dg(\eta_t^*)) - (ch(\eta_t) + dg(\eta_t))] \\
&\mapsto N(0, c^2\sigma_3^2 + d^2\sigma_2^2 + 2cd\sigma_{23}) \text{ in distribution,}
\end{aligned}$$

where $\sigma_2^2 = \text{Var}(g(\eta_1))$, $\sigma_3^2 = \text{Var}(h(\eta_1))$, and $\sigma_{23} = \text{Cov}(h(\eta_1), g(\eta_1))$. \square

Remark: If the bootstrap resample size k' is different from k . Theorem 14 still holds, as long as both k and k' go to ∞ . The rate of growth is irrelevant, but the finite second moment for the hitting time T_Δ is crucial.

3.4 Estimator III

Suppose $\underline{X} \equiv \{X_0, X_1, \dots, X_n\}$ is a realization of an ergodic Markov chain with infinite state space S and transition probability matrix P . The third version of the bootstrap algorithm is as follows:

1) Given n , let $k + 1 = \sup\{T_\Delta^j \leq n\}$ and $\eta_i = (X_{T_\Delta^i}, \dots, X_{T_\Delta^{i+1}-1})$, $i = 1, 2, \dots, k$. Note that unlike in (3.3), here k is a random variable. Let $\hat{\pi}_k(i)$ and $\hat{p}_k(i, j)$ be defined as above.

2) Given \underline{X} , draw a “bootstrap sample” $\underline{X}^* \equiv \{\eta_1^*, \eta_2^*, \dots, \eta_k^*\}$ of size k from the empirical distribution \hat{F}_k .

3) Approximate the sampling distribution of $R(\underline{X}, P) \equiv \sqrt{k}(\hat{P}_k - P)$ by the bootstrap distribution of $R(\underline{X}^*, \hat{P}_k) \equiv \sqrt{k}(\tilde{P}_k - \hat{P}_k)$, where \tilde{P}_k is defined as above.

Even though k is random, it goes to ∞ w.p.1. as $n \mapsto \infty$. Thus, the strong consistency and normality of the bootstrap hold here.

Theorem 15 *With the same notations as above, if there exists $\delta > 0$ such that*

$E_{\Delta} T_{\Delta}^{1+\delta} < \infty$. Then

$$\tilde{\pi}_k(i) - \hat{\pi}_k(i) \mapsto 0 \text{ with probability } 1.$$

Proof. By the definition of the bootstrap estimator, we have

$$\begin{aligned} \tilde{\pi}_k(i) &= \frac{1}{m} \sum_{t=1}^k g(\eta_t^*) \quad (\text{where } m = T_1^* + \dots + T_k^*) \\ &= \frac{k}{m} \frac{1}{k} \sum_{t=1}^k g(\eta_t^*) \\ &= \frac{k}{m} \frac{[\pi_{\Delta} n]}{k} \left[\frac{1}{[\pi_{\Delta} n]} \sum_{t=1}^{[\pi_{\Delta} n]} g(\eta_t^*) \quad (1) + \frac{1}{[\pi_{\Delta} n]} \sum_{t=[\pi_{\Delta} n]+1}^k g(\eta_t^*) \quad (2) \right]. \end{aligned}$$

By Theorem 10 and $E_{\Delta} T_{\Delta}^{1+\delta} < \infty \Rightarrow E_{\Delta} g(\eta_1)^2 < \infty$, we have

a)

$$\begin{aligned} \frac{m}{k} &= \frac{1}{k} \sum_{t=1}^k T_t^* \\ &= \frac{[\pi_{\Delta} n]}{k} \left(\frac{\sum_{t=1}^{[\pi_{\Delta} n]} T_t^*}{[\pi_{\Delta} n]} + \frac{\sum_{t=[\pi_{\Delta} n]+1}^k T_t^*}{[\pi_{\Delta} n]} \right). \end{aligned}$$

The first term converges to $E_{\Delta} T_{\Delta}$ by Theorem 10, the second term converges to 0 as $n \mapsto \infty$, which has been done in Section 2.2.

b) By Theorem 4, we have $k/[\pi_{\Delta} n] \mapsto 1$.

c) By Theorem 10, we have (1) $\mapsto E g(\eta_1)$ with probability 1, and by Theorem 3, we have (2) $\mapsto 0$ with probability 1. \square

Theorem 16 *With the notations given above, if $E_{\Delta} T_{\Delta}^2 < \infty$, then for almost all realizations of the process $\{X_n; n \geq 0\}$, we have*

$$\sqrt{k}(\tilde{\pi}_k(i) - \hat{\pi}_k(i)) \mapsto N(0, \sigma^2) \text{ in distribution,}$$

where σ^2 is the variance which is a function of $Var(T_1)$, $Var(g(\eta_1))$, and $Cov(T_1, g(\eta_1))$.

Proof. Note that

$$\tilde{\pi}_k(i) = \frac{1}{m} \sum_{t=1}^k g(\eta_t^*), \quad \text{where } m = \sum_{t=1}^k T_t^*.$$

Now, consider

$$\begin{aligned} & \sqrt{k}(\tilde{\pi}_k(i) - \hat{\pi}_k(i)) \\ &= \sqrt{k}\left(\frac{1}{m} \sum_{t=1}^k g(\eta_t^*) - \frac{1}{n'} \sum_{t=1}^k g(\eta_t)\right), \end{aligned}$$

where $n' \equiv \sum_{j=1}^k T_j$. In order to prove the convergence of the above equation, we prove the following two results.

$$\begin{aligned} 1) \quad & \sqrt{k}\left(\frac{m}{k} - \frac{n'}{k}\right) \\ &= \sqrt{k}\left[\frac{1}{k}\left(\sum_{j=1}^k T_j^* - \sum_{j=1}^k T_j\right)\right] \\ &= \frac{1}{\sqrt{k}} \sum_{j=1}^k (T_j^* - T_j) \\ &\longmapsto N(0, \sigma_1^2) \text{ in distribution,} \end{aligned}$$

where $\sigma_1^2 = Var(T_1)$.

Since $g(\eta_1^*)$ are i.i.d. and $E_\Delta T_\Delta^2 < \infty \Rightarrow Eg(\eta_1)^2 < \infty$, we have

$$\begin{aligned} 2) \quad & \sqrt{k}\left(\frac{1}{k} \sum_{t=1}^k g(\eta_t^*) - \frac{1}{k} \sum_{t=1}^k g(\eta_t)\right) \\ &\longmapsto N(0, \sigma_2^2) \text{ in distribution,} \end{aligned}$$

where $\sigma_2^2 = \text{Var}(G(\eta_1))$.

By δ – method, we have

$$\sqrt{k}(\tilde{\pi}_k(i) - \hat{\pi}_k(i)) \longmapsto N(0, \sigma^2) \text{ in distribution.} \quad \square$$

We now state the following theorem regarding the asymptotic normality of the distribution of $\sqrt{k}(\tilde{p}_k(i, j) - \hat{p}_k(i, j))$, since the proof is the same as Theorem 14, we do not give it here.

Theorem 17 *With the notations given above, if $E_\Delta T_\Delta^2 < \infty$, then for almost all realizations of the process $\{X_n; n \geq 0\}$, we have*

$$\sqrt{k}(\tilde{p}_k(i, j) - \hat{p}_k(i, j)) \longmapsto N(0, \sigma^2) \text{ in distribution,}$$

where σ^2 is the variance involving $\text{Var}(g(\eta_1))$, $\text{Var}(h(\eta_1))$ and $\text{Cov}(h(\eta_1), g(\eta_1))$.

Remark: We note that for estimators II and III, the bootstrap has been shown to work, whereas for estimator I, we do not have a complete proof yet.

4 LIMIT THEOREMS FOR AN ARRAY OF HARRIS CHAINS

4.1 Central Limit Theorem for an Array of Harris Chains

Let $\{X_n; n \geq 0\}$ be a Harris chain on a measurable state space (S, \mathcal{S}) with transition kernel $P(\cdot, \cdot)$ and reference measure $\varphi(\cdot)$. It is well known that there is a regeneration scheme for such chains in the following sense: there exists an integer $n_0 \geq 1$ for which the skeleton chain $Y_n = \{X_{nn_0}; n \geq 0\}$ can be extended to a chain \tilde{Y}_n on an enlarged space $\tilde{S} = S \cup \{\Delta\}$ with transition kernel \tilde{P} , where Δ is a *recurrent point* for \tilde{Y}_n ; namely $\tilde{P}_x(\tilde{Y}_n = \Delta \text{ for some } n \geq 1) \equiv 1$. Thus, for the Harris chain, without loss of generality, we may assume there exists a recurrence point. For details of this analysis, the reader is referred to Athreya and Ney [7]. We will assume such a recurrent point exists for the whole chapter.

Theorem 18 *The measure*

$$\nu(A) = E_{\Delta} \left(\sum_{j=0}^{T_{\Delta}-1} I_A(X_j) \right),$$

where $T_{\Delta} = \inf\{n : n \geq 1, X_n = \Delta\}$, is an invariant measure for P , namely $\nu P = \nu$. If $\lambda(\cdot)$ is another σ -finite measure satisfying $\lambda P \leq \lambda$, then $\lambda(\cdot) = \lambda(\Delta)\nu(\cdot)$.

The following theorem is a generalization of the classical ergodic theorem for Markov chains. We quote it here without proof, and the reader is referred to Athreya and Ney [7] for the details.

Theorem 19 *Let $E_\Delta T_\Delta < \infty$. Then, for all $f \in L_1(\nu)$, we have*

$$\frac{1}{n} \sum_{j=0}^n f(X_j) - \int f(x) \pi(dx) \longmapsto 0 \text{ almost surely,}$$

where $\pi(dx) = \nu(dx)/\nu(S)$.

The following theorem is a central limit theorem for the functionals

$$\xi_n(f) \equiv \sum_{j=0}^n f(X_j), \quad f \in L_1(\nu).$$

Note that, when $f = I_A$, $\xi_n(A) \equiv \xi_n(I_A)$ contains the number of visits by $\{X_n\}$ to the set A up to time n . The reader is referred to Nummelin [33] for the details.

Theorem 20 *Let $\{X_n\}$ be an irreducible Harris chain with state space (S, \mathcal{S}) and recurrent point Δ with $E_\Delta T_\Delta < \infty$. Let $f \in L_1(\nu)$ be such that*

$$E_\Delta \left(\sum_{j=0}^{T_\Delta-1} f(X_j) \right)^2 < \infty.$$

Then

$$\frac{1}{\sqrt{n}} \sum_{j=0}^n (f(X_j) - \int f(x) \pi(dx)) \longmapsto N(0, \sigma^2(f)) \text{ in distribution,}$$

where

$$\sigma^2(f) = 2 \int f(x)(Tf)(x) \pi(dx) - \int f^2(x) \pi(dx) - \left(\int f(x) \pi(dx) \right)^2,$$

$$\text{and } (Tf)(x) = E_x \left(\sum_{j=0}^{T_\Delta-1} f(X_j) \right).$$

Let $X_n = \{X_{nj}; n \geq 0\}$ be a sequence of Harris chains on a measurable state space (S, \mathcal{S}) with transition kernels $P_n(\cdot, \cdot)$ and reference measures $\varphi_n(\cdot)$. We assume that the existence of recurrent point Δ for the sequence of Harris chains X_n .

Our goal here is to prove a generalization of Theorem 20 to a double array of Harris chains. Let $T_{n\Delta}$ be the first hitting time of the Harris chain X_n to its recurrent point Δ on the state space S . That is,

$$T_{n\Delta} = \begin{cases} \inf \{t \geq 1, X_{nt} = \Delta\}; \\ \infty, & \text{if no such } n \text{ exist.} \end{cases}$$

Let $\nu_n(A) \equiv E_\Delta(\sum_{t=0}^{T_{n\Delta}-1} I_A(X_{nt}))$, we assume $\nu_n(S) \equiv E_\Delta T_{n\Delta} < \infty$ and set $\pi_n(\cdot) \equiv \nu_n(\cdot)/\nu_n(S)$. Define for any $f_n \in L_1(\pi_n)$,

$$\eta_{nj}(f_n) \equiv \sum_{t=T_{n\Delta}^j}^{T_{n\Delta}^{j+1}-1} f_n(X_{nt}).$$

Without loss of generality, we may assume that $E\eta_{n1}(f_n) = 0$.

Theorem 21 *With the notations given above, if*

- 1) $\sup_{x,n} \|P_n^{(\alpha)}(x, \cdot) - \pi_n(\cdot)\| \leq C\rho^\alpha$, for some constant $C > 0$ and $\rho \in (0, 1)$,
where $\|\cdot\|$ is the total variation norm,
- 2) $N_n\pi_{n\Delta} \mapsto \infty$ as $n \mapsto \infty$,
- 3) $\frac{1}{\sigma_n^2} E((\eta_{n1}(f_n))^2; |\eta_{n1}(f_n)| > \epsilon\sigma_n\sqrt{[N_n\pi_{n\Delta}]}) \mapsto 0$ as $n \mapsto \infty$.

where

$$\sigma_n^2 \equiv 2 \int f_n(x)(T_n f_n)(x) \pi_n(dx) - \int f_n^2(x) \pi_n(dx),$$

$$\text{and} \quad (T_n f_n)(x) \equiv E_x\left(\sum_{t=0}^{T_{n\Delta}-1} f_n(X_{nt})\right),$$

then

$$\frac{1}{\sigma_n\sqrt{N_n}} \sum_{j=0}^{N_n} f_n(X_{nj}) \mapsto N(0, 1) \text{ in distribution.}$$

Proof. We decompose $(1/\sigma_n \sqrt{N_n}) \sum_{j=0}^{N_n} f_n(X_{nj})$ as follows:

$$\begin{aligned}
& \frac{1}{\sigma_n \sqrt{N_n}} \sum_{j=0}^{N_n} f_n(X_{nj}) \\
&= \frac{1}{\sigma_n \sqrt{N_n}} \left(\sum_{j=0}^{m_{n\Delta}} \eta_{nj}(f_n) + \sum_{j=T_{n\Delta}^{m_{n\Delta}}}^{N_n} f_n(X_{nj}) \right) \\
&= \frac{\sqrt{[N_n \pi_{n\Delta}]}}{\sqrt{N_n}} \left[\frac{1}{\sigma_n \sqrt{[N_n \pi_{n\Delta}]}} \sum_{j=0}^{[N_n \pi_{n\Delta}]} \eta_{nj}(f_n) \right] \\
&\quad + \frac{1}{\sqrt{N_n}} \sum_{j=[N_n \pi_{n\Delta}]+1}^{m_{n\Delta}} \eta_{nj}(f_n)/\sigma_n + \frac{1}{\sqrt{N_n}} \sum_{j=T_{n\Delta}^{m_{n\Delta}}}^{N_n} f_n(X_{nj})/\sigma_n,
\end{aligned}$$

where $m_{n\Delta}$ is the number of visits to Δ by the n^{th} chain up to time N_n .

Thus, we can prove this theorem by discussing four parts of the above equation separately.

Claim 1. For all $\epsilon > 0$,

$$P\left(\left| \frac{m_{n\Delta}}{N_n} - \pi_{n\Delta} \right| > \epsilon\right) \mapsto 0 \text{ as } n \mapsto \infty.$$

In fact,

$$\begin{aligned}
& E\left(\frac{m_{n\Delta}}{N_n} - \pi_{n\Delta}\right)^2 \\
&= E\left(\frac{1}{N_n} \sum_{j=1}^{N_n} (I(X_{nj} = \Delta) - \pi_{n\Delta})\right)^2 \\
&= \frac{1}{N_n^2} \sum_{l=1}^{N_n} \sum_{k=1}^{N_n} [P(X_{nk} = \Delta, X_{nl} = \Delta) - \pi_{n\Delta} P(X_{nk} = \Delta) \\
&\quad - \pi_{n\Delta} P(X_{nl} = \Delta) + \pi_{n\Delta}^2].
\end{aligned}$$

By assumption 1), the rest of the proof is the same as the proof in Theorem 4.

Claim 2. For all $\epsilon > 0$,

$$P(| \frac{1}{\sqrt{N_n}} \sum_{j=T_n \Delta}^{N_n} f_n(X_{nj}) / \sigma_n | > \epsilon) \mapsto 0 \text{ as } n \mapsto \infty.$$

Let $\epsilon > 0$ be an arbitrary constant, we have

$$\begin{aligned} & P\{ | \sum_{j=T_n \Delta}^{N_n} f_n(X_{nj}) | > \epsilon \sqrt{N_n} \sigma_n \} \\ &= \sum_{m=0}^{N_n} P\{ | \sum_{j=N_n-m}^{N_n} f_n(X_{nj}) | > \epsilon \sqrt{N_n} \sigma_n; \\ & \quad T_{\Delta}^{(m_n \Delta + 1)} > m, T_{\Delta}^{m_n \Delta} = N_n - m \} \\ &= \sum_{m=0}^{N_n} g_n(m) P\{ T_n^{m_n \Delta} = N_n - m \} \text{ (by Markov property)} \\ &\leq \sup_m g_n(m), \end{aligned}$$

where

$$g_n(m) = P_{\Delta}\{ | \sum_{j=0}^m f_n(X_{nj}) | > \epsilon \sqrt{N_n} \sigma_n; T_{n\Delta} > m \}.$$

But

$$\begin{aligned} \sup_m g_n(m) &\leq P_{\Delta}\{ | \sum_{j=0}^{T_{n\Delta}-1} f_n(X_{nj}) | > \epsilon \sqrt{N_n} \sigma_n \} \\ &\leq \frac{E[| \eta_{n1}(f_n) |^2; | \eta_{n1}(f_n) | > \epsilon \sqrt{N_n} \sigma_n]}{\sigma_n^2 \epsilon N_n} \\ &\mapsto 0 \text{ (by assumption 3).} \end{aligned}$$

Claim 3. For all $\epsilon > 0$,

$$P(| \frac{1}{\sigma_n \sqrt{N_n}} \sum_{j=[N_n \pi_{n\Delta}] + 1}^{m_{n\Delta}} \eta_{nj}(f_n) | > \epsilon) \mapsto 0.$$

By Claim 1, we have for any fixed $\epsilon > 0$, an integer n_0 such that

$$P\{ |m_{n\Delta} - [N_n\pi_{n\Delta}]| > \epsilon^3 N_n \} < \epsilon \quad \text{for all } n > n_0.$$

Clearly, for such n ,

$$\begin{aligned} & P\{ \left| \sum_{j=[N_n\pi_{n\Delta}]+1}^{m_{n\Delta}} \eta_{nj}(f_n) \right| > \epsilon \sqrt{N_n \sigma_n} \} \\ & \leq P\{ |m_{n\Delta} - [N_n\pi_{n\Delta}]| > \epsilon^3 N_n \} \\ & \quad + P\{ \max_{|m - [N_n\pi_{n\Delta}]| \leq \epsilon^3 N_n} \left| \sum_{j=[N_n\pi_{n\Delta}]+1}^m \eta_{nj}(f_n) \right| > \epsilon \sqrt{N_n \sigma_n} \} \\ & < \epsilon + 2P\{ \max_{1 \leq m \leq \epsilon^3 N_n} \left| \sum_{j=1}^m \eta_{nj}(f_n) \right| > \epsilon \sqrt{N_n \sigma_n} \} \\ & < \epsilon + \frac{2E(\sum_{j=1}^{[\epsilon^3 N_n]+1} \eta_{nj}(f_n))^2}{\epsilon^2 N_n \sigma_n^2} \quad (\text{by Kolmogorov's inequality}) \\ & = \epsilon + 2\epsilon \\ & = 3\epsilon. \end{aligned}$$

Claim 4.

$$\frac{1}{\sigma_n \sqrt{[N_n\pi_{n\Delta}]}} \sum_{j=0}^{[N_n\pi_{n\Delta}]} \eta_{nj}(f_n) \mapsto N(0,1) \text{ in distribution.}$$

Let

$$Y_{nj} \equiv \frac{\eta_{nj}(f_n)}{\sigma_n \sqrt{[N_n\pi_{n\Delta}]}}.$$

Then, we have for each n

- 1) $\{ Y_{nj}; j = 1, \dots, [N_n\pi_{n\Delta}] \}$ are i.i.d. random variables,
- 2) $EY_{nj} = 0$,
- 3) $\sum_{j=1}^{[N_n\pi_{n\Delta}]} EY_{nj}^2 = \frac{E(\eta_{n1}(f_n))^2}{\sigma_n^2} < \infty$ (by assumption 3).

It is enough to show that $\{Y_{nj}\}$ satisfies the Lindeberg's condition. That is, for any fixed $\epsilon > 0$, we have

$$\sum_{j=1}^{[N_n \pi_n \Delta]} E[Y_{nj}^2 I(|Y_{nj}| > \epsilon)] \longrightarrow 0 \text{ as } n \mapsto \infty.$$

But, the left side of the above equation is equal to

$$\begin{aligned} & \frac{1}{\sigma_n^2} E[(\eta_{n1}(f_n))^2; |\eta_{n1}(f_n)| > \epsilon \sigma_n \sqrt{[N_n \pi_n \Delta]}] \\ & \longrightarrow 0 \text{ as } n \mapsto \infty \text{ (by assumption 3).} \end{aligned}$$

It remains to calculate the variance $E(\eta_{n1}(f_n))^2$. In fact, we have

$$\begin{aligned} & E(\eta_{n1}(f_n))^2 \\ = & E_{\Delta} \left(\sum_{j=0}^{T_{n\Delta}-1} f_n(X_{nj}) \right)^2 \\ = & E_{\Delta} \left(\sum_{i,j=0}^{T_{n\Delta}-1} f_n(X_{ni}) f_n(X_{nj}) \right) \\ = & E_{\Delta} \left(\sum_{i=0}^{T_{n\Delta}-1} f_n^2(X_{ni}) \right. \\ & \left. + 2 E_{\Delta} \left(\sum_{i=0}^{T_{n\Delta}-1} \sum_{j=i+1}^{T_{n\Delta}-1} f_n(X_{ni}) f_n(X_{nj}) \right) \right) \\ = & E_{\Delta} \left(\sum_{i=0}^{T_{n\Delta}-1} f_n^2(X_{ni}) \right. \\ & \left. + 2 E_{\Delta} \left(\sum_{i=0}^{\infty} I(T_{n\Delta} > i) \sum_{j=i+1}^{\infty} f_n(X_{ni}) f_n(X_{nj}) I(T_{n\Delta} > j) \right) \right) \\ = & E_{\Delta} \left(\sum_{i=0}^{T_{n\Delta}-1} f_n^2(X_{ni}) \right. \\ & \left. + 2 \sum_{i=0}^{\infty} E_{\Delta} [(I(T_{n\Delta} > i) f_n(X_{ni})) \cdot E \left(\sum_{j=i+1}^{\infty} f_n(X_{nj}) I(T_{n\Delta} > j) \mid \mathcal{F}_i \right)] \right). \end{aligned}$$

But

$$\begin{aligned}
& E\left[\sum_{j=i+1}^{\infty} f_n(X_{nj})I(T_{n\Delta} > j \mid \mathcal{F}_i)I(T_{n\Delta} > i)\right] \\
&= E_{X_{ni}} \sum_{r=1}^{\infty} f_{nr}(X_{ni})I(T_{n\Delta} > r) \\
&= ((T_n f_n)(X_{ni}) - f_n(X_{ni}))I(T_{n\Delta} > i),
\end{aligned}$$

where $(T_n f_n)(x) \equiv E_x(\sum_{i=0}^{T_n \Delta - 1} f_n(X_{ni}))$.

Hence, the variance

$$\begin{aligned}
& E(\eta_{n1}(f_n))^2 \\
&= E_{\Delta}\left(\sum_{i=0}^{T_n \Delta - 1} f_n^2(X_{ni})\right) \\
&\quad + 2 \sum_{i=0}^{\infty} E_{\Delta}[I(T_{n\Delta} > i)f_n(X_{ni})][((T_n f_n)(X_{ni}) - f_n(X_{ni}))I(T_{n\Delta} > i)] \\
&= 2 \sum_{i=0}^{\infty} E_{\Delta}[(I(T_{n\Delta} > i)f_n(X_{ni}))(T_n f_n)(X_{ni}) - E_{\Delta}\left(\sum_{i=0}^{T_n \Delta - 1} f_n^2(X_{ni})\right)] \\
&= 2 \int f_n(x)(T_n f_n)(x) \pi_n(dx) - \int f_n^2(x) \pi_n(dx). \quad \square
\end{aligned}$$

4.2 Applications and Further Research

Let $\{X_{nj}; j = 1, \dots, N_n, n = 1, \dots\}$ be a sequence of Harris chains with the same assumptions as the previous section. We will show that Theorem 4 is a corollary of Theorem 21. The procedure is as follows:

We define notation first. Let

$$f_{\Delta\Delta}^{(m)} \equiv P_{\Delta}\{X_t \neq \Delta \text{ for } t < m; X_m = \Delta\}.$$

Theorem 22 *Let $\{X_{nj}; j = 1, \dots, N_n\}$ be a sequence of ergodic Markov chains on a finite state space S , and transition probability matrix P_n , such that $P_n(i, j) \mapsto P(i, j)$ for all i, j , where $P(i, j)$ is the transition probability matrix of a Markov chain $\{X_k; k \geq 0\}$ on the same state space S . Let f_n be a sequence of functions on S . If*

$$\sup_{n \geq 1, s \in S} |f_n(s)| < \infty,$$

then

$$\frac{1}{\sqrt{N_n \sigma_n^2(f_n)}} \sum_{j=1}^{N_n} f_n(X_{nj}) \mapsto N(0, 1) \text{ in distribution,}$$

where $\sigma_n^2(f_n)$ is the same as the one in Theorem 21.

Proof. Here, we want to check that with the assumption given above, the three conditions hold in Theorem 21. We prove the following lemma first.

Lemma 6 *With the notations given above, then there exists $\delta > 0$ such that*

$$\sup_n E_{\Delta}[\eta_{n1}^{2+\delta}(f_n)] < \infty.$$

Note that since $\sup_n \sigma_n^2(f_n) < \infty$, Lemma 6 implies assumption 3) in Theorem 21.

Proof. Since the state space S is finite, without loss of generality, we may assume all entries of P_n are positive for all n , and let

$$\begin{aligned} \alpha_n &\equiv \max_{i,j} p_{nij}, \quad \beta_n \equiv \min_{i,j} p_{nij}, \\ \alpha &\equiv \max_{i,j} p_{ij}, \quad \beta \equiv \min_{i,j} p_{ij}. \end{aligned}$$

Then

$$\alpha_n \mapsto \alpha, \text{ and } \beta_n \mapsto \beta.$$

Now, consider

$$\begin{aligned}
f_{\Delta\Delta} &= p_{\Delta\Delta} \leq \alpha, \\
f_{\Delta\Delta}^{(2)} &= \sum_{k \neq \Delta} p_{\Delta k} p_{k\Delta} \leq \alpha \sum_{k \neq \Delta} p_{\Delta k} \\
&= \alpha(1 - p_{\Delta\Delta}) \leq \alpha(1 - \beta), \\
&\vdots \\
f_{\Delta\Delta}^{(m)} &\leq \alpha(1 - \beta)^{m-1}.
\end{aligned}$$

Since $\alpha_n \mapsto \alpha$ and $\beta_n \mapsto \beta$, so, there exists $\epsilon > 0$ and integer n_0 such that

$$f_{n\Delta\Delta}^{(m)} \leq C(1 - \beta - \epsilon)^{m-1}, \quad \text{for all } n \geq n_0.$$

Now, let $\rho > 1$ such that $\rho(1 - \beta - \epsilon) < 1$, then

$$\begin{aligned}
&\sup_{n \geq n_0} E_{\Delta} \rho^{T_{n\Delta}} \\
&= \sup_{n \geq n_0} \sum_{m=1}^{\infty} \rho^m f_{n\Delta\Delta}^{(m)} \\
&\leq \sup_{n \geq n_0} \sum_{m=1}^{\infty} \rho^m C(1 - \beta - \epsilon)^{m-1} \\
&\leq C\rho \frac{1}{1 - \rho(1 - \beta - \epsilon)} \\
&< \infty.
\end{aligned}$$

This implies that there exists $\delta > 0$ such that

$$\sup_{n \geq 1} E_{\Delta} \rho^{T_{n\Delta}} < \infty \implies \sup_{n \geq 1} E_{\Delta} T_{n\Delta}^{2+\delta} < \infty. \quad \square$$

In order to complete the proof, we need to check that

$$1) \quad \sup_{i,n} |P_n^{(\alpha)}(i,j) - \pi_n(j)| \leq C\rho^{\alpha},$$

for some independent constants C and ρ .

$$2) \lim_{n \rightarrow \infty} \inf_{i \neq \Delta} P_n(i, \{\Delta\}) > 0 \Rightarrow N_n \pi_n \Delta \mapsto \infty.$$

For 1), since we have

$$|P_n^{(\alpha)}(i, j) - \pi_n(j)| \leq C \rho_n^\alpha \text{ and } \rho_n = 1 - \alpha_n \text{ where } \alpha_n \mapsto \alpha,$$

so, there exists a constant $\rho \in (0, 1)$ such that

$$\sup_n |P_n^{(\alpha)}(i, j) - \pi_n(j)| \leq C \rho^\alpha.$$

For 2), since Δ is a recurrent point of the Markov chain, so, we have

$$P_n(i, \{\Delta\}) \mapsto P(i, \{\Delta\}) > 0, \text{ as } n \mapsto \infty. \quad \square$$

Now, we give an alternative proof of Theorem 4 which can be regarded as a corollary of Theorem 21. Let us prove the following lemma first.

Lemma 7 *Let X be an ergodic Markov chain with state space (E, \mathcal{E}) and transition probability matrix P . Let g be a measurable function on $(E \times E, \mathcal{E} \otimes \mathcal{E})$ such that $g \in L_1(\mu)$ where μ is $\pi \times \pi$, and π is the invariant measure on E . Then*

$$\begin{aligned} & \frac{1}{n+1} \sum_{m=0}^n g(X_m, X_{m+1}) \\ & \mapsto \int_E \int_E g(x, y) p(x, dy) \pi(dx) \quad \text{with probability 1.} \end{aligned}$$

Proof. Consider $Y_n \equiv (X_n, X_{n+1})$, for $n = 0, 1, \dots$, then Y_n is an ergodic Markov chain with state space $E \times E$, and invariant measure μ . Then, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{m=0}^n g(Y_m) \\ & = \int_{E \times E} g(z) \mu(dz) \\ & = \int_{E \times E} g(x, y) \mu(dx \times dy), \quad \text{where } z = (x, y). \end{aligned}$$

For all $A, B \in E$, $C \equiv (A, B)$, and $\Delta = (i, E)$, we have

$$\begin{aligned}
\mu(C) &= E_{\Delta} \left(\sum_{m=0}^{T_{\Delta}-1} I_C(X_m, X_{m+1}) \right) \\
&= E_i \left(\sum_{m=0}^{T_i-1} I_A(X_m) I_B(X_{m+1}) \right) \\
&= \sum_{m=0}^{\infty} E_i(I_A(X_m) I_B(X_{m+1}) I(T_i > m)) \\
&= \sum_{m=0}^{\infty} E_i[E_i(I_A(X_m) I_B(X_{m+1}) I(T_i > m) \mid X_m)] \\
&= \sum_{m=0}^{\infty} E_i(I_A(X_m) P(X_m, B) I(T_i > m)) \\
&= \int_E I_A(x) p(x, B) \pi(dx) \\
&= \int_A p(x, B) \pi(dx).
\end{aligned}$$

This implies that

$$\begin{aligned}
\mu(dx, dy) &= p(x, dy) \pi(dx) \\
\Rightarrow \int_{E \times E} g(x, y) \mu(dx \times dy) &= \int_E \int_E g(x, y) p(x, dy) \pi(dx). \quad \square
\end{aligned}$$

Theorem 23 *With the notations given above, then for almost all realizations of the process, We have*

$$\sqrt{N_n}(\tilde{P}_n - \hat{P}_n) \mapsto N(\mathcal{Q}, \Sigma_P) \text{ in distribution}$$

as $n \mapsto \infty$ and $N_n \mapsto \infty$, where Σ_P is the same as the one in Theorem 4.

Proof. The maximum likelihood estimator $\hat{P}_n(i, j)$ of $P(i, j)$ is a consistent estimator. That is, $\hat{P}_n(i, j) \mapsto P(i, j)$ for all states i, j in the state space S . For each fixed n ,

the observation $\{X_{nj}; j = i, \dots, N_n\}$ can be regarded as a Markov chain with finite state space S and transition probability matrix $\hat{P}_n(\cdot, \cdot)$. From Theorem 22, we have

$$\sqrt{N_n} \left(\sum_{j=1}^{N_n} \frac{f(X_{nj})}{N_n} - \int f(x) \pi_n(dx) \right) \mapsto N(0, \sigma^2(f)) \text{ in distribution,}$$

and

$$\sqrt{N_n} \left(\sum_{j=1}^{N_n} \frac{g(X_{nj})}{N_n} - \int g(x) \pi_n(dx) \right) \mapsto N(0, \sigma^2(g)) \text{ in distribution.}$$

Then by δ -method, we have

$$\sqrt{N_n} \left(\frac{\sum_{j=0}^{N_n} f(X_{nj})}{\sum_{j=0}^{N_n} g(X_{nj})} - \frac{\int f(x) \pi_n(dx)}{\int g(x) \pi_n(dx)} \right) \mapsto N(0, \sigma^2(f, g)) \text{ in distribution.}$$

Now, we define $Y_n \equiv (X_n, X_{n+1})$, then, Y_n is also a Markov chain, and let

$$f(x, y) = \begin{cases} 1, & \text{if } x = i, y = j; \\ 0, & \text{otherwise,} \end{cases}$$

$$g(x, y) = \begin{cases} 1, & \text{if } x = i; \\ 0, & \text{otherwise.} \end{cases}$$

Then, we have

$$\frac{\sum_{j=0}^{N_n} f(Y_{nj})}{\sum_{j=0}^{N_n} g(Y_{nj})} = \frac{m_{nij}}{m_{ni}},$$

and

$$\frac{\int_E \int_E f(x, y) p(x, dy) \pi_n(dx)}{\int_E \int_E g(x, y) p(x, dy) \pi_n(dx)} = \frac{\hat{P}_n(i, j) \pi_n(i)}{\pi_n(i) \int_E p(x, dy)} = \hat{P}_n(i, j).$$

Hence

$$\begin{aligned} & \sqrt{N_n} \left(\frac{m_{nij}}{m_{ni}} - \hat{P}_n(i, j) \right) \mapsto N(0, \sigma^2(P)) \text{ in distribution} \\ \Rightarrow & \sqrt{N_n} (\tilde{P}_n - \hat{P}_n) \mapsto N(0, \Sigma_p) \text{ in distribution.} \quad \square \end{aligned}$$

The following are further research topics related to this paper.

- 1) The theorems proved in Section 4.2 can only be applied to the finite state Markov chain case. In fact, for a Markov chain with infinite state space, the condition 1) in Theorem 21

$$\sup_{x, \tilde{n}} \| P_n^{(\alpha)}(x, \cdot) - \pi_n(\cdot) \| \leq C \rho^\alpha$$

can not be checked.

- 2) The rate of convergence of the bootstrap estimator is an open question: Is there an Edgeworth type expansion for

$$P\{\sqrt{N_n}(\tilde{P}_n - \hat{P}_n) \leq x\} ?$$

- 3) Bootstrapping the transition kernel $P(\cdot, \cdot)$ of an ergodic Harris chain, is an extension of bootstrapping the transition probability matrix of an ergodic Markov chain with discrete state space. With the Doeblin's condition for the Harris chain, the histogram estimator $\hat{P}_n(\cdot, \cdot)$ of $P(\cdot, \cdot)$ has the asymptotic normality under some regularity conditions. An interesting problem here is to find the asymptotic behavior of the bootstrap estimator.

5 BIBLIOGRAPHY

- [1] M. A. Arcones and E. Gine. *The bootstrap of the mean with arbitrary bootstrap sample size*. Submitted.
- [2] S. Asmussen. *Applied Probability and Queues*. John Wiley and Sons Ltd., New York, 1987.
- [3] K. B. Athreya. *Strong law for the bootstrap*. Statistics and Probability Letters 1 (1983): 147 -150.
- [4] K. B. Athreya. *On the range of recurrent Markov chain*. Statistics and Probability Letters 3 (1985): 143-145.
- [5] K. B. Athreya. *Bootstrap of the mean in the infinite variance case*. Ann. Statist. 15 (1987): 724-731.
- [6] K. B. Athreya. *Bootstrap of the mean in the infinite variance case. I and II*. Technical Report 86-21, 22. Dept. of Statistics, Iowa State University, Ames, Iowa, 1987.
- [7] K. B. Athreya and P. Ney. *A new approach to the limit theory of recurrent Markov chains*. TAMS 245 (1978): 493-501.
- [8] G. J. Babu and K. Singh. *Inference on means using the bootstrap*. Ann. Statist. 11 (1983): 999-1003.
- [9] I. V. Basawa and B. L. S. Prakasa Rao. *Statistical Inference for Stochastic Processes*. Academic Press, London, 1980.
- [10] R. Beran. *Bootstrap method in Statistics*. Jahrb. Math. Ver. 86 (1984): 14-30.
- [11] P. J. Bickel and D. A. Freedman. *Some asymptotic theory for the bootstrap*. Ann. Statist. 9 (1981): 1196-1217.
- [12] P. Billingsley. *Statistical Inference for Markov Processes*. The University of Chicago Press, Chicago, 1961.

- [13] A. Bose. *Edgeworth correction by bootstrap in autoregression*. Ann. Statist. 16 (1988): 1709-1722.
- [14] Y. S. Chow and H. Teicher. *Probability Theory*. 2nd ed. Springer-Verlag, New York, 1988.
- [15] K. L. Chung. *A Course in Probability Theory*. Academic Press, London, 1974.
- [16] C. Derman. *Some asymptotic distribution theory for Markov chains with a denumerable number of states*. Biometrika 43 (1956): 285-294.
- [17] P. Diaconis. *Group Representations in Probability and Statistics*. IMS, Hayward, California, 1988.
- [18] W. Doeblin. *The elements of the general theory of simple Markov chain*. Ann. Sci. Ecole Norm. Sup. III, 57 (1957): 61-111.
- [19] B. Efron. *Bootstrap method: another look at the jackknife*. Ann. Statist. 7 (1979): 1-26.
- [20] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia, 1982.
- [21] W. Feller. *An Introduction to Probability Theory and its Applications*. Vol. II. 2nd ed. Wiley, New York, 1971.
- [22] D. A. Freedman. *Bootstrapping regression models*. Ann. Statist. 9 (1981): 1218-1228.
- [23] D. A. Freedman and S. C. Peters. *Bootstrapping a regression equation: Some empirical results*. J. Am. Statist. Assoc. 79 (1984): 97-106.
- [24] E. Gine and J. Zinn. *Necessary conditions for the bootstrap of the mean*. To appear in Ann. Statist.
- [25] P. Hall. *On the bootstrap and confidence interval*. Ann. Statist. 14 (1986): 1431-1452.
- [26] P. Hall. *Theoretical comparison of bootstrap confidence intervals*. Ann. Statist. 16 (1988): 927-953.
- [27] P. Hall. *Asymptotic properties of the bootstrap for heavy-tailed distributions*. Preprint.

- [28] T. E. Harris. *The existence of stationary measures for certain Markov processes.* p. 113-124 in 3rd Berkeley Symposium. Vol. II. Wiley, Berkeley, California, 1956.
- [29] D. Isaacson and R. Madsen. *Markov Chains.* Wiley, New York, 1976.
- [30] R. J. Kulperger and B. L. S. Prakasa Rao. *Bootstrapping a finite state Markov chain.* To appear in Sankhya (1989).
- [31] J. Neveu. *Mathematical Foundations of the Calculus of Probability.* Holden Day, San Francisco, 1965.
- [32] S. Niemi and E. Nummelin. *Central limit theorem for Markov random walks.* Communications Physico-Mathematicae 54. Societas Scientiarum Fennica, Helsinki, 1982.
- [33] E. Nummelin. *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge University Press, Cambridge, 1984.
- [34] S. Orey. *Limit Theorems for Markov Chain Transition Probabilities.* Van Nostrand, New York, 1971.
- [35] D. Revuz. *Markov Chains.* North Holland, Amsterdam, 1975.
- [36] G. G. Roussas. *Nonparametric estimation in Markov processes.* Ann. Inst. Statist. Math. 21 (1969): 73-87.
- [37] G. G. Roussas. *Nonparametric estimation of the transition distribution function of a Markov process.* Ann. Math. Statist. 40 (1969): 1386-1400.
- [38] K. Singh. *On the asymptotic efficiency of Efron's bootstrap.* Ann. Statist. 9 (1981): 1187-1195.
- [39] C. F. Wu. *Jackknife, bootstrap and other resampling methods in regression analysis.* Ann. Statist. 14 (1986): 1261-1350.